

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: October 26, 2005, 13:04:49 ; Search time 165 Seconds  
(without alignments)  
583.657 Million cell updates/sec

Title: US-09-847-513A-7  
Perfect score: 1301  
Sequence: 1 MKLLILIGSVIALPTFAAG.....NKILFGLIMVAVKSSNA 249

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 366760381 residues

Total number of hits satisfying chosen parameters: 2105600

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 78%  
Listing first 45 summaries

## Database :

A\_Geneseq\_1dbec04:\*  
1: Geneseqp19808:\*  
2: Geneseqp19808:\*  
3: Geneseqp20008:\*  
4: Geneseqp20018:\*  
5: Geneseqp20028:\*  
6: Geneseqp20038:\*  
7: Geneseqp20038:\*  
8: Geneseqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1014.5	78.0	251	5	AAU81215 Proteorho
2	1014.5	78.0	251	5	AAU81193 Proteorho
3	1014.5	78.0	251	5	AAU81216 Proteorho
4	1014.5	78.0	252	8	ADQ82560 Wild type
5	1014.5	78.0	252	8	ADQ82562 Wild type
6	1014.5	78.0	252	8	ADQ82516 Wild type
7	1013.5	77.9	249	8	ADQ82634 Wild type
8	1012.5	77.8	251	5	AAU81208 Proteorho
9	1012.5	77.8	252	5	ADQ82546 Wild type
10	1011.5	77.7	251	5	AAU81214 Proteorho
11	1011.5	77.7	252	8	ADQ82558 Wild type
12	1010.5	77.7	251	5	AAU81209 Proteorho
13	1010.5	77.7	252	8	ADQ82548 Wild type
14	1009.5	77.6	251	5	AAU81211 Proteorho
15	1009.5	77.6	252	8	ADQ82552 Wild type
16	1008.5	77.5	251	5	AAU81213 Proteorho
17	1008.5	77.5	252	8	ADQ82556 Wild type
18	1003.5	77.1	251	5	AAU81192 Proteorho
19	1003.5	77.1	252	8	ADQ82506 Wild type
20	999.5	76.8	252	8	ADQ82688 Mutant ma
21	998.5	76.7	252	8	ADQ82686 Mutant ma
22	997.5	76.7	252	8	ADQ82684 Mutant ma
23	996.5	76.6	252	8	ADQ82676 Mutant ma
24	995.5	76.5	252	8	ADQ82680 Mutant ma
25	995.5	76.5	252	8	ADQ82678 Mutant ma

26	994.5	76.4	252	8	ADQ82674	Adq82674 Mutant ma
27	993.5	76.4	252	8	ADQ82682	Adq82682 Mutant ma
28	794	61.0	254	8	ADQ82662	Adq82662 Wild type
29	794	61.0	254	8	ADQ82660	Adq82660 Wild type
30	776	59.6	254	8	ADQ82658	Adq82658 Wild type
31	764	58.7	254	8	ADQ82656	Adq82656 Wild type
32	737	56.6	230	8	ADQ82666	Adq82666 Wild type
33	177	13.6	259	7	ADQ83525	Adq83525 H. salina
34	177	13.6	262	2	AAQ60691	AAQ60691 Fragment
35	170	13.1	262	2	AAQ60691	AAQ60691 Fragment
36	169.5	13.0	248	2	AAQ60692	AAQ60692 Mature H.
37	169.5	13.0	248	2	AAQ60692	AAQ60692 Mature H.
38	169.5	13.0	248	5	AAQ60692	AAQ60692 Mature H.
39	168.5	13.0	248	3	AAQ60692	AAQ60692 Mature H.
40	165	12.7	214	5	AAQ60692	AAQ60692 Mature H.
41	165	12.7	222	5	AAQ60692	AAQ60692 Mature H.
42	135.5	10.4	250	4	AAQ60692	AAQ60692 Mature H.
43	126	9.7	370	4	AAQ60692	AAQ60692 Mature H.
44	126	9.7	370	6	AAQ60692	AAQ60692 Mature H.
45	117.5	9.0	233	5	AAQ60692	AAQ60692 Mature H.

## ALIGNMENTS

RESULT 1	AAU81215	standard; protein; 251 AA.
ID	AAU81215	standard; protein; 251 AA.
XX	AAU81215	
AC	AAU81215	
DT	07-AUG-2003 (revised)	
DT	09-APR-2002 (first entry)	
XX		
DE	Proteorhodopsin from clone PALE6.	
XX		
KW	Proteorhodopsin, light-driven energy generator; targeted drug delivery;	
KW	biocatalytic reactor; fuel cell; nano-machine; molecular switching;	
KW	data storage; membrane potential; halophilic archaea.	
OS	Eubacteria.	
XX		
FN	WO200183701-A2.	
XX		
PD	08-NOV-2001.	
XX		
PF	02-MAY-2001; 2001WO-US014394.	
XX		
XX	03-MAY-2000; 2000US-0201602P.	
XX	(MONT-) MONTERREY BAY AQUARIUM RES INST.	
XX		
PI	DeJong EF, Beja O;	
XX	WPI; 2002-114151/15.	
DR	N-PSDB; ABR24154.	
XX		
PT	New proteorhodopsin gene for use in a light-driven energy generator.	
PT	comprises an isolated DNA sequence encoding a proteorhodopsin protein.	
XX		
PS	Claim 35; Fig 35; 460p; English.	
XX		
CC	The invention describes a proteorhodopsin gene, comprising an isolated	
CC	DNA sequence for encoding a proteorhodopsin protein. The gene is useful	
CC	in light-driven energy generators. A proteorhodopsin system is useful in	
CC	many industrial and technological applications, for use in targeted drug	
CC	delivery, has primary and secondary energy generators for biocatalytic	
CC	reactors, fuel cells and nano-machines, as well as uses in molecular	
CC	switching or data storage devices. Proteorhodopsin is useful for a	
CC	process to enhance yield or increase the potential of recombinant protein	
CC	production or convert the light-induced membrane potential into cellular	
CC	signals. The system is not restricted to operate in halophilic archaea	
CC	and could therefore be functionally expressed in Escherichia coli and	

Result No.	Score	Query Match	Length	DB	ID	Description
1	1014.5	78.0	251	10	US-09-847-513A-19	Sequence 19, Appl-63
2	1014.5	78.0	251	10	US-09-847-513A-63	Sequence 63, Appl-65
3	1014.5	78.0	251	10	US-09-847-513A-65	Sequence 65, Appl-66
4	1014.5	78.0	252	17	US-10-724-264A-11	Sequence 11, Appl-55
5	1014.5	78.0	252	17	US-10-724-264A-55	Sequence 55, Appl-57
6	1014.5	78.0	252	17	US-10-724-264A-57	Sequence 57, Appl-58
7	1013.5	77.9	249	17	US-10-724-264A-129	Sequence 129, Appl-49
8	1012.5	77.8	251	10	US-09-847-513A-19	Sequence 49, Appl-41
9	1012.5	77.8	252	17	US-10-724-264A-11	Sequence 41, Appl-51
10	1011.5	77.7	251	10	US-09-847-513A-61	Sequence 61, Appl-53
11	1011.5	77.7	252	17	US-10-724-264A-53	Sequence 53, Appl-54

12	1010.5	77.7	251	10	US-09-847-513A-51	Sequence 51, Appl
13	1010.5	77.7	252	17	US-10-724-264A-13	Sequence 43, Appl
14	1009.5	77.6	251	10	US-09-847-513A-55	Sequence 55, Appl
15	1009.5	77.6	252	17	US-10-724-264A-47	Sequence 47, Appl
16	1008.5	77.5	251	10	US-09-847-513A-59	Sequence 59, Appl
17	1008.5	77.5	252	17	US-10-724-264A-51	Sequence 51, Appl
18	1003.5	77.1	251	10	US-09-847-513A-17	Sequence 17, Appl
19	1003.5	77.1	252	17	US-10-724-264A-1	Sequence 1, Appl
20	999.5	76.8	252	17	US-10-724-264A-183	Sequence 183, Appl
21	998.5	76.7	252	17	US-10-724-264A-181	Sequence 181, Appl
22	997.5	76.7	252	17	US-10-724-264A-179	Sequence 179, Appl
23	996.5	76.6	252	17	US-10-724-264A-171	Sequence 171, Appl
24	995.5	76.5	252	17	US-10-724-264A-173	Sequence 173, Appl
25	995.5	76.5	252	17	US-10-724-264A-175	Sequence 175, Appl
26	994.5	76.4	252	17	US-10-724-264A-169	Sequence 169, Appl
27	993.5	76.4	252	17	US-10-724-264A-177	Sequence 177, Appl
28	794	61.0	254	17	US-10-724-264A-155	Sequence 155, Appl
29	794	61.0	254	17	US-10-724-264A-157	Sequence 157, Appl
30	796	59.6	254	17	US-10-724-264A-153	Sequence 153, Appl
31	764	58.7	254	17	US-10-724-264A-151	Sequence 151, Appl
32	737	56.6	230	17	US-10-724-264A-161	Sequence 161, Appl
33	737	56.6	239	18	US-10-510-628-3	Sequence 3, Appl
34	177	13.6	262	16	US-10-688-221-2	Sequence 2, Appl
35	169.5	13.0	228	17	US-10-664-335B-1	Sequence 1, Appl
36	169.5	13.0	248	15	US-10-343-318-1	Sequence 18, Appl
37	165	12.7	214	9	US-09-905-176-18	Sequence 15, Appl
38	165	12.7	222	9	US-09-905-176-15	Sequence 15, Appl
39	157.5	12.1	341	15	US-10-424-569-283641	Sequence 283641, Appl
40	149	11.1	242	15	US-10-425-114-55348	Sequence 55348, Appl
41	146.5	11.3	304	16	US-10-425-115-346442	Sequence 346442, Appl
42	119.5	9.2	233	15	US-10-424-599-174538	Sequence 174538, Appl
43	119.5	9.2	230	16	US-10-767-701-51333	Sequence 51333, Appl
44	117.5	9.0	273	9	US-09-905-176-16	Sequence 16, Appl
45	110	8.5	332	15	US-10-369-493-15522	Sequence 1522, Appl

## ALIGNMENTS

```

1 RESULT 1
2 US-09-847-513A-19
3 Sequence 19, Application US/09847513A
4 Publication No. US20030104375A1
5 GENERAL INFORMATION:
6 APPLICANT: MBARI
7 APPLICANT: Delong, Edward
8 APPLICANT: Beja, Oded
9 TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin
10 FILE REFERENCE: MBA-101
11 CURRENT APPLICATION NUMBER: US/09/847,513A
12 CURRENT FILING DATE: 2001-05-01
13 PRIORITY APPLICATION NUMBER: 60/201,602
14 PRIOR FILING DATE: 2000-05-03
15 NUMBER OF SEQ ID NOS: 65
16 SOFTWARE: PatentIn version 3.0
17 SEQ ID NO 19
18 LENGTH: 251
19 TYPE: PRF
20 ORGANISM: Naturally occurring gamma proteobacterium
21 US-09-847-513A-19

```

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2005, 13:38:51 ; Search time 76 Seconds  
(without alignments)  
1267.149 Million cell updates/sec

Title: US-09-847-513A-7

Perfect score: 1301  
Sequence: 1 MKLLILGSYIALPTFAAG.....NKILFGLIMNVAVKSSNA 249

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 33

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 78%  
Maximum Match 97%  
Listing first 45 summaries

Database: A\_Geneseq\_16Dec04.\*  
1: geneeqp1980s.\*  
2: geneeqp1990s.\*  
3: geneeqp2000s.\*  
4: geneeqp2001s.\*  
5: geneeqp2002s.\*  
6: geneeqp2003as.\*  
7: geneeqp2003bs.\*  
8: geneeqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	96.8	250	5 AAU81202	AAU81202 Proteorho
2	1260	96.8	250	5 AAU81199	AAU81199 Proteorho
3	1260	96.8	251	8 ADQ82528	ADQ82528 Wild type
4	1259	96.8	250	8 ADQ82624	ADQ82624 Wild type
5	1258	96.7	250	5 AAU81196	AAU81196 Proteorho
6	1258	96.7	251	8 ADQ82522	ADQ82522 Wild type
7	1258	96.7	251	8 ADQ82523	ADQ82523 Wild type
8	1257	96.6	250	8 ADQ82620	ADQ82620 Wild type
9	1256	96.5	250	8 ADQ82622	ADQ82622 Wild type
10	1254	96.4	250	5 AAU81197	AAU81197 Proteorho
11	1250	96.1	249	8 ADQ82534	ADQ82534 Wild type
12	1224	94.1	250	8 ADQ82628	ADQ82628 Wild type
13	1098.5	84.4	257	8 ADQ82632	ADQ82632 Wild type
14	1084	83.3	258	8 ADQ82630	ADQ82630 Wild type
15	1075.5	82.7	251	8 ADQ82648	ADQ82648 Wild type
16	1075.5	82.7	251	8 ADQ82652	ADQ82652 Wild type
17	1072.5	82.4	251	8 ADQ82642	ADQ82642 Wild type
18	1071.5	82.4	250	8 ADQ82650	ADQ82650 Wild type
19	1069.5	82.2	251	8 ADQ82644	ADQ82644 Wild type
20	1068.5	82.1	251	8 ADQ82640	ADQ82640 Wild type
21	1065.5	81.9	251	8 ADQ82646	ADQ82646 Wild type
22	1065.5	81.9	251	8 ADQ82646	ADQ82646 Wild type
23	1032.5	79.4	251	5 AAU81210	AAU81210 Proteorho
24	1032.5	79.4	251	5 AAU81212	AAU81212 Proteorho
25	1032.5	79.4	252	8 ADQ82554	ADQ82554 Wild type

26	1032.5	79.4	252	8 ADQ82550	ADQ82550 Wild type
27	1026.5	78.9	251	5 AAU81195	AAU81195 Proteorho
28	1026.5	78.9	251	8 ADQ82638	ADQ82638 Wild type
29	1026.5	78.9	252	8 ADQ82520	ADQ82520 Wild type
30	1019.5	78.4	250	8 ADQ82664	ADQ82664 Wild type
31	1017.5	78.2	249	8 ADQ82636	ADQ82636 Wild type
32	1016.5	78.1	251	5 AAU81194	AAU81194 Proteorho
33	1016.5	78.1	252	8 ADQ82518	ADQ82518 Wild type

## ALIGNMENTS

## RESULT 1

ID AAU81202 standard; protein; 250 AA.

AC AAU81202,

DT 07-AUG-2003 (revised)

DT 09-APR-2002 (first entry)

XX Proteorhodopsin from clone MB40m5.

KW Proteorhodopsin; light-driven energy generator; targeted drug delivery; biocatalytic reactor; fuel cell; nano-machine; molecular switching; data storage; membrane potential; halophilic archaea.

XX Eubacteria.

XX W0200183701-A2.

PD 08-NOV-2001.

XX 02-MAY-2001; 2001MO-US014394.

XX 03-MAY-2000; 2000US-0201602P.

XX (MONT-) MONTEREY BAY AQUARIUM RES INST.

PI DeJong EF, Beja O;

DR WPI; 2002-114151/15.

DR N-PSDB; ABK24141.

XX New proteorhodopsin gene for use in a light-driven energy generator, comprises an isolated DNA sequence encoding a proteorhodopsin protein.

XX Claim 22; Fig 22; 460pp; English.

The invention describes a proteorhodopsin gene, comprising an isolated DNA sequence for encoding a proteorhodopsin protein. The gene is useful in light-driven energy generators. A proteorhodopsin system is useful in many industrial and technological applications, for use in targeted drug delivery, fuel cells and nano-machines, as well as uses in molecular reactors, for data storage devices. Proteorhodopsin is useful for a process to enhance yield or increase the potential of recombinant protein production or convert the light-induced membrane potential into cellular signals. The system is not restricted to operate in halophilic archaea and could therefore be functionally expressed in *Escherichia coli* and other bacteria. It provides for a fast and cheap production method that allows for mass production of functionally active proteorhodopsin. This is the amino acid sequence of a proteorhodopsin protein variant, described in the method of the invention. (Updated on 07-AUG-2003 to correct OS field.)

XX Sequence 250 AA:

Query Match 96.8%; Score 1260; DB 5; Length 250;  
Best Local Similarity 97.2%; Pred. No. 4.2e-133;  
Matches 241; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 26, 2005, 13:38:51 ; Search time 162 Seconds  
(without alignments)  
641.729 Million cell updates/sec

Title: US-09-847-513a-7  
Perfect score: 1301  
Sequence: 1 MKLLILGSLVIALPTFAAG.....NKILFGLIMNVAVKSSNA 249

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 33

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match .78%  
Maximum Match .97%  
Listing first 45 summaries

Database:

Published Applications AA: \*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1260	96.8	250	US-09-847-513a-31	Sequence 31, App1
2	1260	96.8	250	US-09-847-513a-37	Sequence 37, App1
3	1260	96.8	251	US-10-724-264A-23	Sequence 23, App1
4	1259	96.8	250	US-10-724-264A-119	Sequence 119, App1
5	1258	96.7	250	US-09-847-513a-25	Sequence 25, App1
6	1258	96.7	251	US-10-724-264A-17	Sequence 17, App1
7	1258	96.7	251	US-10-724-264A-18	Sequence 18, App1
8	1257	96.6	250	US-10-724-264A-115	Sequence 115, App1
9	1256	96.5	250	US-10-724-264A-117	Sequence 117, App1
10	1254	96.4	250	US-09-847-513a-27	Sequence 27, App1
11	1250	96.1	249	US-10-724-264A-29	Sequence 29, App1

12	1224	94.1	250	17	US-10-724-264A-123	Sequence 123, App
13	1098.5	84.4	257	17	US-10-724-264A-127	Sequence 127, App
14	1084	83.3	258	17	US-10-724-264A-125	Sequence 125, App
15	1075.5	82.7	251	17	US-10-724-264A-143	Sequence 143, App
16	1075.5	82.7	251	17	US-10-724-264A-147	Sequence 147, App
17	1072.5	82.4	251	17	US-10-724-264A-137	Sequence 137, App
18	1071.5	82.4	250	17	US-10-724-264A-145	Sequence 145, App
19	1069.5	82.2	251	17	US-10-724-264A-139	Sequence 139, App
20	1068.5	82.1	251	17	US-10-724-264A-135	Sequence 135, App
21	1066.5	82.0	251	17	US-10-724-264A-149	Sequence 149, App
22	1065.5	81.9	247	17	US-10-724-264A-141	Sequence 141, App
23	1032.5	79.4	251	10	US-09-847-513a-53	Sequence 53, App1
24	1032.5	79.4	251	10	US-09-847-513a-57	Sequence 57, App1
25	1032.5	79.4	252	17	US-10-724-264A-45	Sequence 45, App1
26	1032.5	79.4	252	17	US-10-724-264A-49	Sequence 49, App1
27	1026.5	78.9	251	17	US-09-847-513a-23	Sequence 23, App1
28	1026.5	78.9	251	17	US-10-724-264A-133	Sequence 133, App
29	1026.5	78.9	252	17	US-10-724-264A-115	Sequence 115, App1
30	1019.5	78.4	250	17	US-10-724-264A-159	Sequence 159, App
31	1017.5	78.2	249	17	US-10-724-264A-131	Sequence 131, App
32	1016.5	78.1	251	10	US-09-847-513a-21	Sequence 21, App1
33	1016.5	78.1	252	17	US-10-724-264A-13	Sequence 13, App1

ALIGNMENTS

RESULT 1

US-09-847-513a-31  
; Sequence 31, Application US/09847513A  
; Publication No. US20030104375A1  
; GENERAL INFORMATION:  
; APPLICANT: MBARI  
; APPLICANT: DeLong, Edward  
; APPLICANT: Beja, Oded  
; TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin  
; FILE REFERENCE: M8A-101  
; CURRENT APPLICATION NUMBER: US/09/847, 513A  
; PRIOR FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,602  
; PRIORITY DATE: 2000-05-03  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 31  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Naturally occurring gamma proteobacterium  
; US-09-847-513a-31

Query Match 96.8%; Score 1260; DB 10; Length 250;  
Best Local Similarity 97.6%; Pred. No. 6.3e-117;  
Matches 242; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY	2	KLILIGSVIALPTFAAGGDDLDASDTGVSFWLVTAALIASTVFVEEDRVSAKRTS	61
DB	3	KLILIGSVIALPTFAAGGDDLDASDTGVSFWLVTAALIASTVFVEEDRVSAKRTS	62
QY	62	LTVSGVLTGIAFMHYMVRGVTETGDSPTVPRYIDLVLPILICEFYILIAATVAVG	121
DB	63	LTVSGVLTGIAFMHYMVRGVTETGDSPTVPRYIDLVLPILICEFYILIAATVAVG	122
QY	122	SLFVKLVGSLVLMVLFVGYMGEGAIMAPAFITICLAVMYIYELMAGEKSACTASPA	181
DB	123	SLFVKLVGSLVLMVLFVGYMGEGAIMAPAFITICLAVMYIYELMAGEKSACTASPS	182
QY	182	VQSAVNTMYITIIIFGMATVVGFTGYLMGDGGSALNLTINYLADEFVNKILFGLITWV	241
DB	183	VQSAVNTMYITIIIFGMATVVGFTGYLMGDGGSALNLTINYLADEFVNKILFGLITWV	242
QY	242	AVKSSNA 249	
DB	243	AVKSSNA 250	



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 09:23:31 ; Search time 169 Seconds

(without alignments)  
754.484 Million cell updates/sec

Title: US-09-847-513A-7

Perfect score: 1301  
Sequence: 1 MKLLILGCVIALPTFAAG.....NKILFGLIIMVAVKSSNA 249

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: 1: uniprot\_03;  
2: uniprot\_trembl;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1301	100.0	249 1 PRG_PRB01	Q91P74 gamma-prote
2	1296	99.6	250 2 Q6U4G7	Q6J1G7 uncultured
3	1296	99.6	251 2 Q9AFB6	Q9AFB6 uncultured
4	1293	99.4	250 2 Q84BD7	Q84BD7 uncultured
5	1290	99.2	250 2 Q84C28	Q84C28 uncultured
6	1289	99.1	250 2 Q83TU0	Q83TU0 uncultured
7	1289	99.1	250 2 Q84BE3	Q84BE3 uncultured
8	1288	99.0	251 2 Q9AFB5	Q9AFB5 uncultured
9	1287	98.9	250 2 Q84BE4	Q84BE4 uncultured
10	1286	98.8	250 2 Q83UH6	Q83UH6 uncultured
11	1286	98.8	250 2 Q84BE8	Q84BE8 uncultured
12	1285	98.8	250 2 Q84BE2	Q84BE2 uncultured
13	1285	98.8	250 2 Q84BE9	Q84BE9 uncultured
14	1284	98.7	251 2 Q9AFB1	Q9AFB1 uncultured
15	1283	98.6	250 2 Q84BD9	Q84BD9 uncultured
16	1283	98.6	250 2 Q84BE5	Q84BE5 uncultured
17	1283	98.6	251 2 Q9AFB9	Q9AFB9 uncultured
18	1282	98.5	250 2 Q83TG9	Q83TG9 uncultured
19	1282	98.5	250 2 Q84BE0	Q84BE0 uncultured
20	1281	98.5	249 2 Q9AFB2	Q9AFB2 uncultured
21	1280	98.4	250 2 Q84BE1	Q84BE1 uncultured
22	1280	98.4	250 2 Q84BE5	Q84BE5 uncultured
23	1279	98.3	250 2 Q84C29	Q84C29 uncultured
24	1277	98.2	250 2 Q84C27	Q84C27 uncultured
25	1275	98.0	250 2 Q84BE7	Q84BE7 uncultured
26	1274	97.9	251 2 Q9AFB8	Q9AFB8 uncultured
27	1273	97.8	249 2 Q84C31	Q84C31 uncultured
28	1273	97.8	250 2 Q84C30	Q84C30 uncultured
29	1273	97.8	252 2 Q9AFB9	Q9AFB9 uncultured
30	1272	97.8	250 2 Q84BD8	Q84BD8 uncultured
31	1270	97.6	251 2 Q9AFB7	Q9AFB7 uncultured

#### ALIGNMENTS

RESULT 1	ID	PRG_PRB01	STANDARD	PRT	249 AA
AC	Q9P7F4	10-OCT-2003 (Rel. 42, Created)			Q84C23 uncultured
DT	10-OCT-2003 (Rel. 42, Last sequence update)				Q84C32 uncultured
DT	10-OCT-2003 (Rel. 43, Last annotation update)				Q99GB3 uncultured
DE	29-MAR-2004 (Rel. 43, Last annotation update)				Q9AFG0 uncultured
DE	Green-light absorbing proteorhodopsin precursor (GPR).				Q9AFB3 uncultured
OS	Bacteria; Proteobacteria; Gammaproteobacteria; environmental samples.				Q84C24 uncultured
OX	NCBI_TaxId=135804;				Q84C25 uncultured
RN	(1)				Q84C26 uncultured
RP	SEQUENCE FROM N.A.				Q84C27 uncultured
RX	MEDLINE=20446260; PubMed=10988064; DOI=10.1126/science.289.5486.1902;				Q84C28 uncultured
RA	Beja O., Araavind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.,				Q84C29 uncultured
RA	Jovanovich S.B., Gates C.M., Feldman R.A., Spudich J.L., Spudich E.N.,				Q84C30 uncultured
RA	DeLong E.P.,				Q84C31 uncultured
RT	"Bacterial rhodopsin: evidence for a new type of phototrophy in the				Q84C32 uncultured
RT	sea."				Q84C33 uncultured
RL	Science 289:1902-1906 (2000).				Q84C34 uncultured
RN	(2)				Q84C35 uncultured
RP	CHARACTERIZATION OF PROTON RELEASE CYCLE, AND MUTAGENESIS OF CYS-107;				Q84C36 uncultured
RP	CYS-156 AND CYS-175.				Q84C37 uncultured
RX	PubMed=11943070; DOI=10.1186/1472-6793-2-5;				Q84C38 uncultured
RT	Krebs R.A., Alexiev U., Partha R., Devita A., Braiman M.S.,				Q84C39 uncultured
RT	"Detection of fast light-activated H <sup>+</sup> release and M intermediate				Q84C40 uncultured
RL	formation from proteorhodopsin."				Q84C41 uncultured
RN	BMC Physiol. 2:5-5(2002).				Q84C42 uncultured
RP	(3)				Q84C43 uncultured
RX	CHARACTERIZATION OF PHOTOCHEMICAL CYCLE, AND MUTAGENESIS OF ASP-97 AND				Q84C44 uncultured
RP	GLU-108.				Q84C45 uncultured
RX	MEDLINE=21965984; PubMed=11969395; DOI=10.1021/bi025563x;				Q84C46 uncultured
RA	Dioumaev A.K., Brown L.S., Shih J., Spudich E.N., Spudich J.L.,				Q84C47 uncultured
RA	Lanyi J.K.,				Q84C48 uncultured
RT	"Proton transfers in the photochemical reaction cycle of				Q84C49 uncultured
RT	proteorhodopsin."				Q84C50 uncultured
RL	Biochemistry 41:5348-5358(2002).				Q84C51 uncultured
RN	(4)				Q84C52 uncultured
RP	CHARACTERIZATION OF PHOTOCHEMICAL CYCLE.				Q84C53 uncultured
RX	MEDLINE=22434728; PubMed=12547799;				Q84C54 uncultured
RA	Varo G., Brown L.S., Lakatos M., Lanyi J.K.,				Q84C55 uncultured
RT	"Characterization of the photochemical reaction cycle of				Q84C56 uncultured
RT	proteorhodopsin."				Q84C57 uncultured
RL	Biophys. J. 84:1202-1207(2003).				Q84C58 uncultured
RN	(5)				Q84C59 uncultured
RP	COMPARISON WITH PHOTOCHEMICAL CYCLE OF BLUE PROTEORHODOPSIN.				Q84C60 uncultured
RX	PubMed=12821661; DOI=10.1074/jbc.M305716200;				Q84C61 uncultured
RA	Wang W.W., Sineshchekov O.A., Spudich E.N., Spudich J.L.,				Q84C62 uncultured
RT	"Spectroscopic and photochemical characterization of a deep ocean				Q84C63 uncultured
RT	proteorhodopsin."				Q84C64 uncultured
RL	J. Biol. Chem. 278:33985-33991(2003).				Q84C65 uncultured
RN	(6)				Q84C66 uncultured
RP	PRELIMINARY CHARACTERIZATION OF PUMP VECTORIALITY.				Q84C67 uncultured
RX	MEDLINE=22195742; PubMed=12206764; DOI=10.1016/S0022-2836(02)00696-4;				Q84C68 uncultured

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using SW model

Run on: October 25, 2005, 09:38:38 ; Search time 39 Seconds  
(without alignments)  
614.307 Million cell updates/sec

Title: US-09-847-513a-7

Perfect score: 1301

Sequence: 1 MKLLILGSVIALPTFAAG.....NKILFGLIIMVAVKSSNA 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187.5	14.4	239	2	SS5300
2	177	13.6	262	2	RAHSB
3	177	13.6	262	2	H84300
4	169.5	13.0	237	1	T44947
5	169.5	13.0	237	2	H84327
6	169.5	13.0	249	4	T44821
7	169.5	13.0	259	4	T44820
8	168	12.9	258	2	T50676
9	168	12.9	260	2	A34178
10	163.5	12.6	259	2	S14731
11	163	12.5	236	2	S55297
12	153.5	11.8	250	2	T50677
13	149.5	11.5	250	2	SS1206
14	141.5	10.9	209	2	B47686
15	140	10.8	204	2	C47686
16	137.5	10.6	211	2	A47686
17	134	10.3	291	2	A35002
18	129.5	10.0	282	2	T43840
19	127	9.8	284	2	S29888
20	118.5	9.1	274	1	A26161
21	118.5	9.1	274	2	G84178
22	117.5	9.0	254	2	T43843
23	116.5	9.0	276	2	T48842
24	113.5	8.7	276	2	T48843
25	112	8.6	239	2	A56808
26	109	8.4	332	2	S18488
27	109	8.4	512	2	G90399
28	106	8.1	261	2	AF2201
29	105.5	8.1	407	2	AF1716

30	105	8.1	298	2	A83892	hypothetical prote
31	105	8.1	361	2	C84246	ribose ABC transpo
32	104	8.0	239	1	S09277	sensory rhodopsin
33	104	8.0	239	2	F84318	sensory rhodopsin
34	102.5	7.9	412	2	A83604	probable MFS trans
35	102.5	7.9	426	2	UC5086	polytopic cytoplas
36	102.5	7.9	453	2	H83790	sodium-dependent t
37	102	7.8	292	2	T49829	related to YRO2 pr
38	102	7.8	490	2	S77201	hypothetical prote
39	101.5	7.8	407	2	AC1346	antipporter protein
40	101.5	7.8	525	2	S26022	cytochrome-c oxida
41	101.5	7.8	606	2	S66857	probable membrane
42	101	7.8	297	2	T43838	halorhodopsin [slm
43	100	7.7	820	2	G75251	cytochrome-c oxida
44	99.5	7.6	247	2	S29889	sensory rhodopsin
45	99.5	7.6	315	2	A83435	procheme IX farne

## ALIGNMENTS

### RESULT 1

SS5300  
Sensory rhodopsin II - Natronobacterium pharaonis  
N:Alternate names: phoborhodopsin  
C:Species: Natronobacterium pharaonis  
C>Date: 14-Oct-1995 #sequence-revision 12-Apr-1996 #text-change 09-Jul-2004  
C/Accession: SS5300; S55301  
R:Seidel, R.; Schaff, B.; Gautel, M.; Kleine, K.; Oesterhalt, D.; Engelhard, M.  
Proc. Natl. Acad. Sci. U.S.A. 92, 1036-1040, 1995  
A>Title: The primary structure of sensory rhodopsin II: a member of an additional retinal  
A:Reference number: S55296; MUID:95224074; PMID:7708770  
A:Accession: S55300  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-239 <SEI>  
A:Cross-References: UNIPROT:P42196; EMBL:Z35086; NID:G510868; PID:G510870  
A:Experimental source: strain SPI  
A:Accession: S55301  
A:Molecule type: protein  
A:Residues: 2-31 <SBM>  
A:Experimental source: strain SPI  
A:Accession: S55301  
A:Gene: SRII

C:Superfamily: bacteriorhodopsin  
C:Keywords: chromoprotein, photoreceptor, retinal, transmembrane protein  
F:8-25/Domain: transmembrane #status predicted <TM1>  
F:33-55/Domain: transmembrane #status predicted <TM2>  
F:71-90/Domain: transmembrane #status predicted <TM3>  
F:99-115/Domain: transmembrane #status predicted <TM4>  
F:127-145/Domain: transmembrane #status predicted <TM5>  
F:158-181/Domain: transmembrane #status predicted <TM6>  
F:193-215/Domain: transmembrane #status predicted <TM7>  
F:205/Bifiding site: retinal (Lys) (covalent) #status predicted

Query Match 14.4% Score 187.5; DB 2; Length 239;  
Best Local Similarity 30.8% Pred. No. 5.2e-09;  
Matches 69; Conservative 42; Mismatches 84; Indels 29; Gaps 13;

QY	33	FVLVTALLASTV-FFVERDRVSAKMTSLTVSGLVGTIAFWHYMR-GV-WIETGDS	89
DB	8	FVLGAIGMLVGLTAPLAWAGRDGSGERRVYTLVG-IGIAAVAVVVALGVGWPAER	66
QY	90	PLVFF--RYIDMLTVPLICEFYLLAATVAGSLFFKLLVGSVLMVFGMGAGIMA	147
DB	67	-TVFAPRYIDMLTVPLVFGVLAGLDSREFGV--ITLNTVVMKA---GFAQAV	118
QY	148	----AMPFIIGCTLAVMYIELMAGEGKSACNTASPAVQSAVNTMTYIIIFGMAIYVG	203
DB	119	PEIERVAFMGCAVAFLLVYL-VGPMTEGASQSSGSIKYVLRNLTVLMAIYF-	176
QY	204	YFTGLMGDGSAL-----NLNLYNLADFYVKKILFGLIIMVVA	242

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 09:20:28 ; Search time 43 seconds  
(without alignments)  
432.270 Million cell updates/sec

Title: US-09-847-513A-7  
Perfect score: 1301  
Sequence: 1 MKLLILGIVIALPTFAGG.....NKILFGLIWNVAKSSNA 249

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues  
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/aa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/aa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/aa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/aa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	177	13.6	262	1 US-08-313-553-2	Sequence 2, Appl
2	177	13.6	262	3 US-08-767-993-2	Sequence 2, Appl
3	177	13.6	559	1 US-08-313-553-15	Sequence 15, Appl
4	177	13.6	559	3 US-08-767-993-15	Sequence 15, Appl
5	169.5	13.0	248	1 US-08-313-553-3	Sequence 2, Appl
6	169.5	13.0	248	1 US-08-303-651-2	Sequence 2, Appl
7	169.5	13.0	248	3 US-08-767-993-3	Sequence 2, Appl
8	102.5	7.9	466	4 US-09-252-991A-27479	Sequence 27479, A
9	102	7.8	405	4 US-09-603-208A-262	Sequence 262, App
10	99	7.6	465	4 US-09-489-039A-13505	Sequence 13505, A
11	98.5	7.6	465	4 US-09-602-787A-530	Sequence 530, App
12	97.5	7.5	355	4 US-09-248-796A-14258	Sequence 14258, A
13	95	7.3	537	4 US-09-252-991A-32657	Sequence 32657, A
14	94	7.2	439	4 US-09-107-532A-4123	Sequence 4123, Ap
15	93.5	7.2	439	4 US-09-252-991A-28709	Sequence 28709, A
16	93	7.1	327	4 US-09-489-039A-8641	Sequence 8641, Ap
17	92.5	7.1	336	4 US-09-252-991A-20404	Sequence 20404, A
18	92.5	7.1	452	3 US-09-134-001C-3935	Sequence 3935, Ap
19	92.5	7.1	465	4 US-09-198-452A-524	Sequence 524, App
20	92.5	7.1	664	4 US-09-328-352-7056	Sequence 7056, Ap
21	92	7.1	394	4 US-09-107-532A-3292	Sequence 3292, Ap
22	92	7.1	476	4 US-09-134-001C-3778	Sequence 3778, Ap
23	91.5	7.0	489	4 US-09-328-352-5088	Sequence 5088, Ap
24	91	7.0	249	4 US-09-282-218A-17	Sequence 17, Appl
25	90.5	7.0	373	4 US-09-603-208A-174	Sequence 174, Appl
26	90.5	7.0	381	4 US-09-248-796A-20097	Sequence 20097, A
27	90	6.9	554	3 US-09-134-001C-5109	Sequence 5109, Ap

28	89.5	6.9	293	4 US-09-134-000C-6254	Sequence 6254, Ap
29	89	6.8	300	4 US-09-393-634-17	Sequence 17, Appl
30	88.5	6.8	465	4 US-09-438-185A-488	Sequence 488, Appl
31	88.5	6.8	719	4 US-09-328-352-6274	Sequence 6274, Appl
32	88	6.8	452	4 US-09-543-681A-6544	Sequence 6544, Ap
33	88	6.8	554	4 US-09-440-236-2634	Sequence 2634, Ap
34	87.5	6.7	264	3 US-09-134-001C-3780	Sequence 3780, Ap
35	87.5	6.7	355	4 US-09-489-039A-9302	Sequence 9302, Ap
36	87.5	6.7	536	4 US-09-328-352-4689	Sequence 4689, Ap
37	87.5	6.7	657	4 US-09-252-991A-27682	Sequence 27682, A
38	87	6.7	455	4 US-09-543-681A-7043	Sequence 7043, Ap
39	87	6.7	502	4 US-09-489-039A-13185	Sequence 13185, A
40	86.5	6.6	349	3 US-09-134-001C-4004	Sequence 4004, Ap
41	86.5	6.6	395	4 US-09-328-352-5819	Sequence 5819, Ap
42	86.5	6.6	940	4 US-09-328-352-8165	Sequence 8165, Ap
43	86	6.6	246	4 US-09-710-279-900	Sequence 900, Appl
44	86	6.6	246	4 US-09-710-279-1956	Sequence 1956, Appl
45	86	6.6	445	4 US-09-602-777A-312	Sequence 312, Appl

## ALIGNMENTS

```
RESULT 1
US-08-313-553-2
; Sequence 2, Application US/08313553
; Patent No. 5641650
; GENERAL INFORMATION:
; APPLICANT: TURNER, George J.
; APPLICANT: BETLACH, Mary C.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES
; TITLE OF INVENTION: IN HALOBACTERIA
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,553
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,662
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-57669/WHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 396-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-313-553-2
; Query Match 13.6%; Score 177; DB 1; Length 262;
; Best local similarity 27.8%; Pred. No. 3.4e-10;
; Matches 69; Conservative 40; Mismatches 101; Indels 38; Gaps 13;
OY 13 LPTFAAGGDLDAADYGVSWF---LVYTAALASTVFFVERDRVS-AKWKTSITVSG 67
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 09:20:28 : Search time 902 Seconds  
(without alignments)  
115.255 Million cell updates/sec

Title: US-09-847-513a-7  
Perfect score: 1301  
Sequence: 1 MKLLILGSLVIALPFAAGS.....NKILFGLIWNVAVKSSNA 249

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues  
Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1301	100.0	249	10	US-09-847-513a-7
2	1301	100.0	249	17	US-10-724-264a-3
3	1296	99.6	250	10	US-09-847-513a-5
4	1296	99.6	250	10	US-09-847-513a-45
5	1296	99.6	251	17	US-10-724-264a-37
6	1294	99.5	249	17	US-10-724-264a-165
7	1293	99.4	249	17	US-10-724-264a-167
8	1293	99.4	250	17	US-10-724-264a-101
9	1292	99.3	249	17	US-10-724-264a-163
10	1292	99.3	250	10	US-09-847-513a-11
11	1291	99.2	250	10	US-09-847-513a-33

12	1290	99.2	250	17	US-10-724-264a-111	Sequence 111, App
13	1289	99.1	250	17	US-10-724-264a-65	Sequence 65, App1
14	1289	99.1	250	17	US-10-724-264a-69	Sequence 69, App1
15	1289	99.1	250	17	US-10-724-264a-77	Sequence 77, App1
16	1289	99.1	250	17	US-10-724-264a-83	Sequence 83, App1
17	1289	99.1	250	17	US-10-724-264a-85	Sequence 85, App1
18	1289	99.1	250	17	US-10-724-264a-87	Sequence 87, App1
19	1288	99.0	250	17	US-09-847-513a-47	Sequence 47, App1
20	1288	99.0	251	10	US-10-724-264a-39	Sequence 39, App1
21	1287	98.9	250	17	US-10-724-264a-61	Sequence 61, App1
22	1286	98.8	250	17	US-10-724-264a-73	Sequence 73, App1
23	1286	98.8	250	17	US-10-724-264a-91	Sequence 91, App1
24	1286	98.8	250	17	US-10-724-264a-59	Sequence 59, App1
25	1285	98.8	250	17	US-10-724-264a-89	Sequence 89, App1
26	1285	98.8	250	17	US-09-847-513a-35	Sequence 35, App1
27	1284	98.7	251	10	US-10-724-264a-27	Sequence 27, App1
28	1284	98.7	250	17	US-09-847-513a-39	Sequence 39, App1
29	1283	98.6	250	17	US-10-724-264a-71	Sequence 71, App1
30	1283	98.6	250	17	US-10-724-264a-97	Sequence 97, App1
31	1283	98.6	251	17	US-10-724-264a-31	Sequence 31, App1
32	1283	98.5	250	17	US-10-724-264a-63	Sequence 63, App1
33	1282	98.5	250	17	US-10-724-264a-75	Sequence 75, App1
34	1282	98.5	250	17	US-10-724-264a-95	Sequence 95, App1
35	1282	98.5	249	17	US-10-724-264a-25	Sequence 25, App1
36	1281	98.5	250	17	US-10-724-264a-79	Sequence 79, App1
37	1280	98.4	250	17	US-10-724-264a-93	Sequence 93, App1
38	1280	98.4	250	17	US-10-724-264a-109	Sequence 109, App
39	1279	98.3	250	17	US-10-724-264a-103	Sequence 103, App
40	1277	98.2	250	17	US-10-724-264a-113	Sequence 113, App
41	1277	98.2	250	17	US-10-724-264a-67	Sequence 67, App1
42	1275	98.0	250	10	US-09-847-513a-41	Sequence 41, App1
43	1274	97.9	251	10	US-10-724-264a-33	Sequence 33, App1
44	1274	97.9	251	10	US-10-724-264a-105	Sequence 105, App
45	1273	97.8	249	17		

## ALIGNMENTS

RESULT 1  
US-09-847-513a-7  
; Sequence 7, Appli  
; Publication No. US20030104375A1  
; GENERAL INFORMATION:  
; APPLICANT: NEARI  
; APPLICANT: Delong, Edward  
; APPLICANT: Beja, Oded  
; TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin  
; FILE REFERENCE: MBA-101  
; CURRENT APPLICATION NUMBER: US/09/847,513A  
; PRIOR FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,602  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Naturally occurring gamma proteobacterium  
US-09-847-513a-7

Query Match 100.0%; Score 1301; DB 10; Length 249;  
Best Local Similarity 100.0%; Pred. No. 5.2e-121;  
Matches 249; Conservative 0; Indels 0; Gaps 0;  
Db 61 SLTVSGLVGTGIAFMHYMYMRGWIETGDSPTVFRIYDMLITVPLLICFYLIAAATNVA 120  
Oy 1 MKLLILGSLVIALPFAAGCDLDSYDYGVSFMLVTAALASTVFFVERDRYSAKKKT 60  
Oy 61 SLTVSGLVGTGIAFMHYMYMRGWIETGDSPTVFRIYDMLITVPLLICFYLIAAATNVA 120  
Db 61 SLTVSGLVGTGIAFMHYMYMRGWIETGDSPTVFRIYDMLITVPLLICFYLIAAATNVA 120

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 19, 2005, 08:08:47 ; Search time 3301 Seconds

(without alignments)  
3655.059 Million cell updates/sec

Title: US-09-847-513A-7

Perfect score: 1301  
Sequence: 1 MKLLILIGSVIALPTFAAGS.....NKILFGLIMVAVKESNA 249

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2.1/USPRO.spool.p/US09847513/runat.19102005.090520.24129/app.query.fasta.1.391  
-DB=GenMbl -OPMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human0.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09847513 @CGN 1.1 5600 @runat.19102005.090520.24129 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

## Database :

GenMbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1301	100.0	128758	1	AF279106 Unculture
2	1296	99.6	751	1	AY601905 Unculture
3	1296	99.6	753	1	AF349993 Unculture
4	1293	99.4	751	1	AY210919 Unculture

5	1290	99.2	751	1	AY250718 Unculture
6	1289	99.1	750	1	AY210910 Unculture
7	1289	99.1	751	1	AY210901 Unculture
8	1289	99.1	751	1	AY210903 Unculture
9	1289	99.1	751	1	AY210907 Unculture
10	1289	99.1	751	1	AY210911 Unculture
11	1289	99.1	751	1	AY210912 Unculture
12	1288	99.0	753	1	AF349994 Unculture
13	1287	98.9	750	1	AY210909 Unculture
14	1286	98.8	751	1	AY210899 Unculture
15	1286	98.8	751	1	AY210905 Unculture
16	1286	98.8	751	1	AY210914 Unculture
17	1285	98.8	751	1	AY210898 Unculture
18	1285	98.8	751	1	AY210913 Unculture
19	1284	98.7	753	1	AF349988 Unculture
20	1283	98.6	751	1	AY210904 Unculture
21	1283	98.6	751	1	AY210917 Unculture
22	1283	98.6	753	1	AF349990 Unculture
23	1282	98.5	751	1	AY210900 Unculture
24	1282	98.5	751	1	AY210906 Unculture
25	1282	98.5	751	1	AY210916 Unculture
26	1281	98.5	748	1	AF349987 Unculture
27	1280	98.4	750	1	AY210908 Unculture
28	1280	98.4	751	1	AY210915 Unculture
29	1279	98.3	751	1	AY250717 Unculture
30	1277	98.2	751	1	AY250719 Unculture
31	1275	98.0	751	1	AY210902 Unculture
32	1274	97.9	753	1	AF349991 Unculture
33	1273	97.8	748	1	AY250715 Unculture
34	1273	97.8	748	1	AY598756 Unculture
35	1273	97.8	751	1	AY250716 Unculture
36	1273	97.8	751	1	AF349977 Unculture
37	1272	97.8	751	1	AY210918 Unculture
38	1270	97.6	753	1	AF349992 Unculture
39	1268	97.5	751	1	AY250723 Unculture
40	1267	97.4	747	1	AY250714 Unculture
41	1267	97.4	753	1	AF349985 Unculture
42	1267	97.4	753	1	AF349978 Unculture
43	1262	97.0	757	1	AF349976 Unculture
44	1260	96.8	753	1	AF349986 Unculture
45	1259	96.8	751	1	AY250722 Unculture

## ALIGNMENTS

RESULT 1	AF279106/c	128758 bp	DNA	linear	BCT 29-APR-2004
LOCUS	AF279106	Uncultured marine gamma proteobacterium	EBAC31A08	BAC sequence.	
DEFINITION	AF279106	Uncultured marine gamma proteobacterium	EBAC31A08	BAC sequence.	
ACCESSION	AF279106	GI:34112904			
VERSION	AF279106.2	GI:34112904			
KEYWORDS					
SOURCE					
ORGANISM		uncultured marine gamma proteobacterium	EBAC31A08		
REFERENCE		Bacteria; Proteobacteria; Gammaproteobacteria; environmental			
AUTHORS		1 (bases 1 to 128758)			
TITLE		Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A.,			
JOURNAL		Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A.,			
MEDLINE		Spudich, J. L., Spudich, E. N. and Delong, E. F.			
PUBMED		Bacterial rhodopsin: evidence for a new type of phototrophy in the			
REFERENCE		sea			
AUTHORS		Science 289 (5486), 1902-1906 (2000)			
TITLE		2 (bases 1 to 128758)			
JOURNAL		Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A.,			
MEDLINE		Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A. and			
PUBMED		DeLong, E. F.			
REFERENCE		Direct Submission			
AUTHORS		Submitted (15-JUN-2000) RAD, Monterey Bay Aquarium Research			
TITLE		Institute, 7700 Sandholdt Road, Moss Landing, CA 95039-0628, USA			
JOURNAL		3 (bases 1 to 128758)			

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 19, 2005, 08:05:27 ; Search time 452 Seconds  
(without alignments)  
3261.096 Million cell updates/sec

Title: US-09-847-513A-7

Perfect score: 1301  
Sequence: 1 MLLILGLSVIALPTFAGG.....NKLLFGLIINVAKSSNA 249

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ .p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool.p/US09847513/runat.19102005.090520.24121/app.query.fasta.1.391  
-DB=N Genesegq.16dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALL=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORML=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09847513 @CGN 1.1 708 @runat.19102005.090520.24121 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Genesegq.16dec04:\*  
1: genesegq1980s:\*  
2: genesegq1990s:\*  
3: genesegq2000s:\*  
4: genesegq2001as:\*  
5: genesegq2001bs:\*  
6: genesegq2002as:\*  
7: genesegq2002bs:\*  
8: genesegq2003as:\*  
9: genesegq2003bs:\*  
10: genesegq2003cs:\*  
11: genesegq2003ds:\*  
12: genesegq2004as:\*  
13: genesegq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1301	100.0	747	6	ABK24126
2	1301	100.0	750	13	ADQ82509
3	1301	100.0	105184	6	ABK24122
4	1296	99.6	750	6	ABK24125
5	1296	99.6	750	6	ABK24145

6	1296	99.6	753	13	ADQ82543	Adq82543 Wild type
7	1294	99.5	750	13	ADQ82671	Adq82671 Mutant ma
8	1293	99.4	750	13	ADQ82673	Adq82673 Mutant ma
9	1293	99.4	751	13	ADQ82607	Adq82607 Wild type
10	1292	99.3	750	6	ABK24128	Abk24128 DNA encod
11	1292	99.3	750	13	ADQ82669	Adq82669 Mutant ma
12	1291	99.2	750	6	ABK24139	Abk24139 DNA encod
13	1290	99.2	751	13	ADQ82617	Adq82617 Wild type
14	1289	99.1	750	13	ADQ82589	Adq82589 Wild type
15	1289	99.1	751	13	ADQ82591	Adq82591 Wild type
16	1289	99.1	751	13	ADQ82575	Adq82575 Wild type
17	1289	99.1	751	13	ADQ82583	Adq82583 Wild type
18	1289	99.1	751	13	ADQ82593	Adq82593 Wild type
19	1289	99.1	751	13	ADQ82571	Adq82571 Wild type
20	1288	99.0	750	6	ABK24146	Abk24146 DNA encod
21	1288	99.0	753	13	ADQ82545	Adq82545 Wild type
22	1287	98.9	750	13	ADQ82587	Adq82587 Wild type
23	1286	98.8	751	13	ADQ82597	Adq82597 Wild type
24	1286	98.8	751	13	ADQ82567	Adq82567 Wild type
25	1286	98.8	751	13	ADQ82579	Adq82579 Wild type
26	1285	98.8	751	13	ADQ82595	Adq82595 Wild type
27	1285	98.7	751	13	ADQ82565	Adq82565 Wild type
28	1284	98.7	750	6	ABK24140	Abk24140 DNA encod
29	1284	98.7	753	13	ADQ82533	Adq82533 Wild type
30	1283	98.6	750	6	ABK24142	Abk24142 DNA encod
31	1283	98.6	751	13	ADQ82577	Adq82577 Wild type
32	1283	98.6	751	13	ADQ82603	Adq82603 Wild type
33	1283	98.6	753	13	ADQ82537	Adq82537 Wild type
34	1282	98.5	751	13	ADQ82601	Adq82601 Wild type
35	1282	98.5	751	13	ADQ82569	Adq82569 Wild type
36	1282	98.5	751	13	ADQ82581	Adq82581 Wild type
37	1281	98.5	748	13	ADQ82531	Adq82531 Wild type
38	1280	98.4	750	13	ADQ82585	Adq82585 Wild type
39	1280	98.4	751	13	ADQ82599	Adq82599 Wild type
40	1279	98.3	751	13	ADQ82615	Adq82615 Wild type
41	1277	98.2	751	13	ADQ82609	Adq82609 Wild type
42	1277	98.2	751	13	ADQ82619	Adq82619 Wild type
43	1275	98.0	751	13	ADQ82573	Adq82573 Wild type
44	1274	97.9	750	6	ABK24143	Abk24143 DNA encod
45	1274	97.9	753	13	ADQ82539	Adq82539 Wild type

#### ALIGNMENTS

RESULT 1	ABK24126	standard; DNA; 747 BP.
ID	ABK24126;	
AC	ABK24126;	
XX		
XX		
DT	07-AUG-2003 (revised)	
DT	09-APR-2002 (first entry)	
XX		
DE	DNA encoding Proteorhodopsin from clone EBAC31A8 #2.	
XX		
KW	Proteorhodopsin; light-driven energy generator; targeted drug delivery;	
KW	biocatalytic reactor; fuel cell; nano-machine; molecular switching;	
KW	data storage; membrane potential; halophilic archaea; ds.	
XX		
OS	Eubacteria.	
XX		
PN	WO200183701-A2.	
XX		
PD	08-NOV-2001.	
XX		
PF	02-MAY-2001; 2001WO-US014394.	
XX		
PR	03-MAY-2000; 2000US-0201602P.	
XX		
PA	(MONT-) MONTEREY BAY AQUARIUM RES INST.	
XX		
PI	DeJong EF, Beja O;	
XX		

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: October 19, 2005, 16:18:24 ; Search time 171 Seconds  
(without alignments)  
2382.649 Million cell updates/sec

Title: US-09-847-513A-7

Perfect score: 1301  
Sequence: 1 MKLLLGSIYALPTFAAG.....NKILFGILINVAKSSNA 249

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_p2n.model -DEV=x1p  
-Q=/cgn2\_1/USPTO.spool\_P/US09847513/runat\_19102005\_090521\_24155/app\_query.fasta\_1.391  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdd  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09847513 @CIGN\_1\_1\_105 @runat\_19102005\_090521\_24155 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NRG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PTCUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	177	13.6	1254	1 US-08-313-553-1	Sequence 1, Appl
2	177	13.6	1254	3 US-08-767-993-1	Sequence 1, Appl
3	177	13.6	2147	1 US-08-313-553-14	Sequence 14, Appl
4	177	13.6	2147	3 US-08-767-993-14	Sequence 14, Appl
5	103.5	8.0	1830121	4 US-09-557-884-1	Sequence 1, Appl
6	103.5	7.9	1401	4 US-09-643-990A-1	Sequence 1, Appl
7	102.5	7.9	1401	4 US-09-252-991A-10908	Sequence 10908, A
8	102.5	7.9	1440	4 US-09-252-991A-11041	Sequence 11041, A
9	102	7.8	1338	4 US-09-603-208A-261	Sequence 261, App
10	99	7.6	501	4 US-09-252-991A-11149	Sequence 11149, A
11	99	7.6	1368	4 US-09-602-787A-529	Sequence 529, App
12	98.5	7.6	1383	4 US-09-489-039A-6334	Sequence 6334, Ap

13	97.5	7.5	1068	4 US-09-248-796A-155	Sequence 155, App
14	97.5	7.5	42325	4 US-08-311-731A-131	Sequence 131, Appl
15	96	7.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl
16	96	7.3	1470	4 US-09-328-352-962	Sequence 1, Appl
17	95.5	7.3	1259	2 US-08-672-814D-12	Sequence 12, Appl
18	95	7.3	1259	3 US-09-333-696-12	Sequence 12, Appl
19	95	7.3	1754	4 US-09-252-991A-16086	Sequence 16086, A
20	95	7.3	1878	4 US-09-252-991A-16516	Sequence 16516, A
21	95	7.2	1432	4 US-09-107-532A-469	Sequence 469, App
22	94	7.2	1330	4 US-09-252-991A-12138	Sequence 12138, A
23	93.5	7.2	8536	4 US-08-956-171E-278	Sequence 278, App
24	93.5	7.2	8536	4 US-08-781-986A-278	Sequence 278, App
25	93.5	7.1	984	4 US-09-489-039A-1470	Sequence 1470, Ap
26	93	7.1	1776	4 US-09-252-991A-15987	Sequence 15987, A
27	93	7.1	3164	1 US-08-188-228-49	Sequence 49, Appl
28	93	7.1	3164	1 US-08-332-643-43	Sequence 43, Appl
29	93	7.1	3164	1 US-08-332-638-49	Sequence 49, Appl
30	93	7.1	4000	4 US-09-016-434-1434	Sequence 1434, Ap
31	93	7.1	4000	4 US-09-023-655-1452	Sequence 1452, Ap
32	93	7.1	23439	4 US-08-956-171E-38	Sequence 38, Appl
33	93	7.1	23439	4 US-08-781-986A-38	Sequence 38, Appl
34	93	7.1	1011	4 US-09-252-991A-3833	Sequence 3833, Ap
35	92.5	7.1	1290	4 US-09-252-991A-3792	Sequence 3792, Ap
36	92.5	7.1	1359	3 US-09-134-001C-1098	Sequence 1098, Ap
37	92.5	7.1	2156	4 US-09-328-352-2930	Sequence 2930, Ap
38	92.5	7.1	2156	4 US-09-710-279-4428	Sequence 4428, Ap
39	92.5	7.1	2742	4 US-09-252-991A-3678	Sequence 3678, Ap
40	92.5	7.1	1230025	4 US-09-198-452A-1	Sequence 1, Appl
41	92.5	7.1	1830121	4 US-09-557-884-1	Sequence 1, Appl
42	92.5	7.1	1830121	4 US-09-643-990A-1	Sequence 1, Appl
43	92.5	7.1	1180	4 US-09-710-279-3291	Sequence 3291, Appl
44	92	7.1	1428	3 US-09-134-001C-941	Sequence 941, App
45	92	7.1			

# ALIGNMENTS

RESULT 1  
US-08-313-553-1  
Sequence 1, Application US/08313553  
Patent No. 5641650  
GENERAL INFORMATION:  
APPLICANT: TURNER, George J.  
INVENTOR: BERTACH, Mary C.  
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES  
NUMBER OF INVENTIONS: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Walter H. Dreger  
STREET: 4 Embarradero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,553  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,662  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-57669/WHD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n.model

Run on: October 19, 2005, 17:38:44 ; Search time 693 Seconds

(without alignments)  
2519.111 Million cell updates/sec

Title: US-09-847-513A-7

Perfect score: 1301

Sequence: 1 MKLLIGSVIALPTFAAG.....NKILFGLIINWVAVKSSNA 249

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Xgapop 10.0	Xgapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 8766186 segs, 3505510206 residues

Total number of hits satisfying chosen parameters: 17532372

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xld  
-Q=cgnt2\_1/USFTO.spool\_p/US09847513/runat\_19102005\_090522\_24184/app\_query.fasta\_1.391  
-DB=published.Applications\_NA -QFWT=fastcap -SUFFIX=p2n.rmpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HRAPOST=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09847513 @CGN 1 1 723 @runat\_19102005\_090522\_24184  
-NCDU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*

```
1: /cgnt2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgnt2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq.*
3: /cgnt2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgnt2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgnt2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgnt2_6/ptodata/2/pubpna/PTCUS_PUBCOMB.seq.*
7: /cgnt2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgnt2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgnt2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgnt2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgnt2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgnt2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgnt2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgnt2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
15: /cgnt2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
16: /cgnt2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
17: /cgnt2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
18: /cgnt2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
19: /cgnt2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
20: /cgnt2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
21: /cgnt2_6/ptodata/2/pubpna/US10J_PUBCOMB.seq.*
22: /cgnt2_6/ptodata/2/pubpna/US10K_PUBCOMB.seq.*
23: /cgnt2_6/ptodata/2/pubpna/US10L_PUBCOMB.seq.*
24: /cgnt2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
25: /cgnt2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
26: /cgnt2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1301	100.0	747	10	US-09-847-513A-6
2	1301	100.0	750	21	US-10-724-264A-4
3	1301	100.0	105184	10	US-09-847-513A-1
4	1296	99.6	750	10	US-09-847-513A-4
5	1296	99.6	753	21	US-10-724-264A-38
6	1294	99.5	750	21	US-10-724-264A-166
7	1293	99.4	750	21	US-10-724-264A-168
8	1293	99.4	751	21	US-10-724-264A-102
9	1292	99.3	750	10	US-09-847-513A-10
10	1292	99.3	750	21	US-10-724-264A-164
11	1291	99.2	750	10	US-09-847-513A-32
12	1290	99.2	751	21	US-10-724-264A-112
13	1289	99.1	751	21	US-10-724-264A-84
14	1289	99.1	751	21	US-10-724-264A-66
15	1289	99.1	751	21	US-10-724-264A-70
16	1289	99.1	751	21	US-10-724-264A-78
17	1289	99.1	751	21	US-10-724-264A-82
18	1289	99.1	751	21	US-10-724-264A-86
19	1289	99.1	751	21	US-10-724-264A-88
20	1288	99.0	750	10	US-09-847-513A-46
21	1288	99.0	753	21	US-10-724-264A-40
22	1287	98.9	750	21	US-10-724-264A-82
23	1286	98.8	751	21	US-10-724-264A-74
24	1286	98.8	751	21	US-10-724-264A-62
25	1285	98.8	751	21	US-10-724-264A-92
26	1285	98.8	751	21	US-10-724-264A-60
27	1285	98.8	751	21	US-10-724-264A-90
28	1284	98.7	750	10	US-09-847-513A-34
29	1284	98.7	753	21	US-10-724-264A-28
30	1283	98.6	750	10	US-09-847-513A-38
31	1283	98.6	751	21	US-10-724-264A-72
32	1283	98.6	751	21	US-10-724-264A-98
33	1283	98.6	751	21	US-10-724-264A-32
34	1282	98.5	751	21	US-10-724-264A-64
35	1282	98.5	751	21	US-10-724-264A-76
36	1282	98.5	751	21	US-10-724-264A-96
37	1281	98.5	748	21	US-10-724-264A-26
38	1280	98.4	750	21	US-10-724-264A-80
39	1280	98.4	751	21	US-10-724-264A-94
40	1279	98.3	751	21	US-10-724-264A-110
41	1277	98.2	751	21	US-10-724-264A-104
42	1277	98.0	751	21	US-10-724-264A-114
43	1275	98.0	751	21	US-10-724-264A-68
44	1274	97.9	750	10	US-09-847-513A-40
45	1274	97.9	753	21	US-10-724-264A-34

#### ALIGNMENTS

```
RESULT 1
US-09-847-513A-6
; Sequence 6, Application US/09847513A
; Publication No. US20030104375A1
; GENERAL INFORMATION:
; APPLICANT: MBARI
; APPLICANT: DeLong, Edward
; APPLICANT: Beja, Oded
; TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin
; FILE REFERENCE: MBA-101
; CURRENT APPLICATION NUMBER: US/09/847,513A
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,602
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.0
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 19, 2005, 13:47:19 ; Search time 3084 Seconds  
(without alignments)  
3073.282 Million cell updates/sec

Title: US-09-847-513a-7

Perfect score: 1301  
Sequence: 1 MKLLILISVIALPTFPAAG.....NKILFGLIMVAVKSSNA 249

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlp  
-Q=/gpn2.1/USPTO-epool.p/US09847513/runat.19102005\_090521\_24141/app\_query.fasta.1.391  
-DB=EST -QFMT=fastap -SUPFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=Pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847513\_QCGN\_1.1\_5180@runat.19102005\_090521\_24141 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=30 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hic:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gsa1:  
9: gb\_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	173.5	13.3	809	2	AW342219 GHESST105
2	164	12.6	627	1	AJ639139
3	155	11.9	752	1	AJ639344
4	148.5	11.4	715	7	CF259014
5	142	10.9	661	1	AJ637035
6	142	10.9	702	1	AJ639451
7	140	10.8	631	1	AJ639452
8	140	10.8	668	1	AJ638560
9	138	10.6	660	8	BZ891866

10	132.5	10.2	607	1	AJ638781
11	130	10.0	872	6	CA764330
12	128.5	9.9	931	7	CO008926
13	126.5	9.7	780	7	CO012533
14	126.5	9.7	871	7	CO011377
15	126.5	9.7	879	7	CO014037
16	126.5	9.7	883	7	CO025968
17	126.5	9.7	912	7	CO004549
18	125.5	9.6	609	1	AJ638428
19	121	9.3	720	7	CO012879
20	121	9.3	828	7	CO034658
21	121	9.3	875	7	CO004790
22	119.5	9.2	717	7	CK448874
23	119	9.1	713	7	CK446419
24	116	8.9	1662	9	CG756213
25	115.5	8.9	603	1	AJ636093
26	115.5	8.9	672	1	AJ638198
27	115.5	8.9	687	1	AJ636129
28	115.5	8.9	716	7	CK446337
29	115	8.8	701	7	CV131522
30	114.5	8.8	592	7	CK448876
31	114	8.8	582	7	CF643569
32	112.5	8.6	703	2	BF251110
33	112.5	8.6	715	7	CO012144
34	112	8.6	612	7	CF639423
35	111.5	8.6	941	2	BF683746
36	110.5	8.5	952	5	BO890217
37	110	8.5	1507	9	CG756860
38	108.5	8.3	585	8	BZ305206
39	108	8.3	413	7	CF642219
40	108	8.3	882	7	CF699316
41	107.5	8.3	888	6	CD458359
42	107.5	8.3	947	6	CD458575
43	104.5	8.0	969	9	AG129346
44	104	8.0	565	6	CD489118
45	104	8.0	646	7	CF641333

#### ALIGNMENTS

RESULT 1  
AW342219/c  
LOCUS  
DEINITION  
AW342219  
VERSION  
AW342219  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AW342219 809 bp mRNA linear EST 01-JAN-2001  
GHESST105 Guillardia theta Lambda Zap II cDNA Library Guillardia  
theta cDNA clone Bsc272\_7b, mRNA sequence.  
AW342219  
GI:12000594  
Guillardia theta  
Guillardia theta  
Bukariyota; Cryptophyta; Cryptomonadaceae; Guillardia.  
1 (bases 1 to 809)  
Fraunholz, M., Duebel, J., Maier, J., Zauner, S. and Maier, U.-G.  
EST Database of the cryptomonad alga: Guillardia theta  
Unpublished (2000)  
Contact: Maier, U.-G.  
Department of Cell Biology and Applied Botany  
Philipps-University Marburg  
Karl-von-Friesch-Strasse, D-35043 Marburg, Germany  
Tel: ++49 6421 282 2057  
Fax: ++49 6421 282 1543  
Email: maier@uni-marburg.de.

#### FEATURES

source  
1..809  
location/Qualifiers  
/organism="Guillardia theta"  
/mol\_type="mRNA"  
/strain="CMP327"  
/db\_xref="taxon:55529"  
/clone="Bsc272\_7b"  
/lab\_host="SOLR"  
/clone\_lib="Guillardia theta Lambda Zap II cDNA Library"  
/note="Vector: Lambda Zap II; Site\_1: EcoRI"

ORIGIN

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 14:04:19 / Search time 3035 Seconds

(without alignments)  
10069.025 Million cell updates/sec

Title: US-09-847-513a-6

Perfect score: 747

Sequence: 1 atgaattactactgatact.....ttaagaatctctaatgct 747

Scoring table: IDENTITY NTC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813366 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1:  
1: gb\_ba:\*  
2: gb\_ba:\*  
3: gb\_ba:\*  
4: gb\_ba:\*  
5: gb\_ba:\*  
6: gb\_ba:\*  
7: gb\_ba:\*  
8: gb\_ba:\*  
9: gb\_ba:\*  
10: gb\_ba:\*  
11: gb\_ba:\*  
12: gb\_ba:\*  
13: gb\_ba:\*  
14: gb\_ba:\*  
15: gb\_ba:\*  
16: gb\_ba:\*  
17: gb\_ba:\*  
18: gb\_ba:\*  
19: gb\_ba:\*  
20: gb\_ba:\*  
21: gb\_ba:\*  
22: gb\_ba:\*  
23: gb\_ba:\*  
24: gb\_ba:\*  
25: gb\_ba:\*  
26: gb\_ba:\*  
27: gb\_ba:\*  
28: gb\_ba:\*  
29: gb\_ba:\*  
30: gb\_ba:\*  
31: gb\_ba:\*  
32: gb\_ba:\*  
33: gb\_ba:\*  
34: gb\_ba:\*  
35: gb\_ba:\*  
36: gb\_ba:\*  
37: gb\_ba:\*  
38: gb\_ba:\*  
39: gb\_ba:\*  
40: gb\_ba:\*  
41: gb\_ba:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	747	100.0	105184	1	AF279106
2	740.8	99.2	750	1	AY210909
3	739.2	99.0	751	1	AY210913
4	739.2	99.0	751	1	AY210919
5	737.6	98.7	750	1	AY210910
6	736	98.5	750	1	AY210908
7	734.4	98.3	751	1	AY210918
8	734.4	98.3	751	1	AY210914
9	734.4	98.3	751	1	AY210917
10	734.2	98.3	748	1	AF349987
11	731.4	97.9	748	1	AY250715
12	731.2	97.9	751	1	AY250716
13	728	97.5	753	1	AF349990
14	728	97.5	753	1	AF349993
15	724.8	97.0	751	1	AY210917
16	713.6	95.5	753	1	AF349986
17	705.6	94.5	753	1	AF349982
18	705.6	94.5	753	1	AF349985
19	704	94.2	753	1	AF349991
20	700.8	93.4	757	1	AF349977
21	697.6	93.4	753	1	AF349994
22	696	93.2	753	1	AF349978
23	694.4	93.0	757	1	AF349976
24	694.4	93.0	753	1	AF349983
25	689.6	92.3	748	1	AF349989
26	688	92.1	753	1	AF349984
27	688	92.1	751	1	AY210911
28	667.2	88.2	751	1	AY210905
29	659.2	88.2	751	1	AY210914
30	659.2	88.2	751	1	AY250719
31	657.6	88.0	751	1	AY250718
32	654.4	87.6	751	1	AY210915
33	652.8	87.4	751	1	AY210915
34	651.2	87.2	751	1	AY250720
35	648	86.7	751	1	AY210901
36	648	86.7	751	1	AY210903
37	648	86.7	751	1	AY210904
38	648	86.7	751	1	AY210906
39	646.4	86.5	751	1	AY210898
40	646.4	86.5	751	1	AY210900
41	644.8	86.3	751	1	AY210899
42	644.8	86.3	751	1	AY210902
43	644.8	86.3	751	1	AY210912
44	644.8	86.3	751	1	AY210916
45	644.8	86.3	751	1	AY210916

## ALIGNMENTS

RESULT 1  
AF279106/c 105184 bp DNA linear BCT 23-OCT-2000  
LOCUS AF279106/c 105184 bp DNA linear BCT 23-OCT-2000  
DEFINITION Uncultured proteobacterium EBAC31A08 clone BAC EBAC31A08, complete  
ACCESSION AF279106  
VERSION AF279106  
KEYWORDS  
SOURCE uncultured proteobacterium EBAC31A08  
ORGANISM Bacteria; Proteobacteria; environmental samples.  
REFERENCE 1 (bases 1 to 105184)  
AUTHORS Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A.,  
Spudich,U.L., Spudich,B.N. and DeLong,E.F.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 14:04:19 / Search time 67 Seconds  
(without alignments)  
4921.093 Million cell updates/sec

Title: US-09-847-513A-6

Perfect score: 747  
Sequence: 1 atgaatacttactgatatc.....ttaagaactcttactgct 747

Scoring table: IDENTITY\_NNC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents\_NA:  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.2	5.8	1497	US-09-220-132-94	Sequence 94, Appl
2	42.2	5.6	6794	US-09-491-356C-2	Sequence 2, Appl
3	42	5.6	6793	US-08-470-203-56	Sequence 56, Appl
4	42	5.6	6793	US-08-471-770-56	Sequence 56, Appl
5	42	5.6	6793	US-08-468-059-56	Sequence 56, Appl
6	42	5.6	6793	US-09-109-916-56	Sequence 56, Appl
7	42	5.6	6793	US-09-886-156-56	Sequence 56, Appl
8	42	5.6	6793	US-09-886-149-56	Sequence 56, Appl
9	42	5.6	6793	US-09-886-150-56	Sequence 56, Appl
10	42	5.6	6793	US-09-328-332-2534	Sequence 56, Appl
11	41.6	5.6	1188	US-09-545-528D-1	Sequence 56, Appl
12	41.6	5.6	580073	US-09-134-001C-2190	Sequence 2190, Ap
13	39.8	5.3	774	US-09-491-356C-7	Sequence 7, Appl
14	39.8	5.3	6558	US-07-991-867B-25	Sequence 25, Appl
15	38.8	5.2	1395	US-08-107-755A-25	Sequence 25, Appl
16	38.8	5.2	1395	US-08-544-132-25	Sequence 25, Appl
17	38.8	5.2	1395	US-08-544-132-25	Sequence 25, Appl
18	38.8	5.2	1395	US-09-370-861A-25	Sequence 25, Appl
19	38.8	5.2	6766	US-08-107-755A-1	Sequence 1, Appl
20	38.8	5.2	8457	US-07-991-867B-1	Sequence 1, Appl
21	38.8	5.2	8457	US-08-544-132-1	Sequence 1, Appl
22	38.8	5.2	8457	US-09-370-861A-1	Sequence 1, Appl
23	38.6	5.1	1159	US-09-181-585-1	Sequence 1, Appl
24	38.4	5.1	5511	US-08-928-361B-2	Sequence 2, Appl
25	38.4	5.1	5511	US-09-588-995A-2	Sequence 2, Appl
26	38.4	5.1	7334	US-08-928-361B-1	Sequence 1, Appl
27	38.4	5.1	7334	US-09-588-995A-1	Sequence 1, Appl

C	28	38.2	5.1	319608	4	US-09-539-333D-1	Sequence 1, Appl
C	29	38.2	5.1	319608	4	US-09-679-409-1	Sequence 1, Appl
C	30	38	5.1	1284	4	US-09-134-001C-2306	Sequence 2306, Ap
C	31	37.8	5.1	580073	4	US-08-545-528D-1	Sequence 1, Appl
C	32	37.6	5.0	9873	4	US-09-328-352-1360	Sequence 1360, Ap
C	33	37.2	5.0	397	4	US-09-253-691-3	Sequence 3, Appl
C	34	37	5.0	1037	4	US-09-181-581-3	Sequence 3, Appl
C	35	37	5.0	1471	4	US-09-181-585-2	Sequence 2, Appl
C	36	36.8	4.9	1086	1	US-08-415-751-27	Sequence 27, Appl
C	37	36.8	4.9	1086	1	US-08-415-751-28	Sequence 28, Appl
C	38	36.8	4.9	5163	3	US-08-700-651-1	Sequence 1, Appl
C	39	36.8	4.9	5163	3	US-08-700-651-1	Sequence 1, Appl
C	40	36.8	4.9	5163	4	US-09-588-995A-4	Sequence 4, Appl
C	41	36.8	4.9	5318	3	US-08-700-651-2	Sequence 2, Appl
C	42	36.8	4.9	5318	3	US-08-928-361B-3	Sequence 3, Appl
C	43	36.8	4.9	5318	4	US-09-588-995A-3	Sequence 3, Appl
C	44	36.8	4.9	1664976	4	US-08-916-421B-1	Sequence 1, Appl
C	45	36.6	4.9	265	4	US-09-491-356C-14	Sequence 14, Appl

## ALIGNMENTS

```
RESULT 1
US-09-220-132-94
Sequence 94, Application US/09220132
Patent No. 6506607
GENERAL INFORMATION:
APPLICANT: Shy[an, Andrew W.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANC
FILE REFERENCE: 07334-074001
CURRENT APPLICATION NUMBER: US/09/220,132
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: US 60/079,303
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: US 60/068,821
PRIOR FILING DATE: 1997-12-24
NUMBER OF SEQ ID NOS: 191
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 94
LENGTH: 1497
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) - (1497)
OTHER INFORMATION: n = A,T,C or G
US-09-220-132-94
Query Match
Best Local Similarity 5.8%, Score 43.2; DB 4; Length 1497;
Matches 62; Conservative 85; Mismatches 98; Indels 1; Gaps 1;
112 CCGCTTATTATGACATCTACTGATTTCTTCTGTTAAAGATAGATTCGCAAA 171
326 SYKMMKKKKKMMMAAAMWTTTAAAMKGRAMYKRAAARRSMWTCRRSMMM 385
172 TGGAAACATCATTAAGTATGATCTGCTGTTACTGATTTGCTTTCGCAATCATG 231
386 WYTWMMMAAMWTTTAAAMWTTTAAAMKGRAMYKRAAARRSMWTCRRSMMM 445
232 TACATGAGAGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 291
446 SMARGRMTGMMMAAMWTTTAAAMKGRAMYKRAAARRSMWTCRRSMMM 505
292 TCGTTTAAACAGTCTCTATTAATGATGATGATGATGATGATGATGATGATGAT 351
506 WATWMMMAAMWTTTAAAMWTTTAAAMKGRAMYKRAAARRSMWTCRRSMMM 564
352 AATGTT 357
565 AATATT 570
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 9, 2004, 16:16:59 / Search time 341 Seconds

(without alignments)  
8069.424 Million cell updates/sec

Title: US-09-847-513a-6

Perfect score: 747

Sequence: 1 atgaatctactctgatac.....ttaaagaatcttcaatgc 747

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	747	100.0	747	US-09-847-513a-6	Sequence 6, Appli
2	747	100.0	105184	US-09-847-513a-1	Sequence 1, Appli
3	744	99.6	750	US-09-847-513a-4	Sequence 32, Appli
4	739.2	99.0	750	US-09-847-513a-32	Sequence 32, Appli
5	737.6	98.7	750	US-09-847-513a-10	Sequence 10, Appli
6	728	97.5	750	US-09-847-513a-38	Sequence 38, Appli
7	728	97.5	750	US-09-847-513a-44	Sequence 44, Appli
8	713.6	95.5	750	US-09-847-513a-34	Sequence 34, Appli
9	705.6	94.5	750	US-09-847-513a-30	Sequence 30, Appli
10	705.6	94.5	750	US-09-847-513a-42	Sequence 42, Appli
11	704	94.2	750	US-09-847-513a-28	Sequence 28, Appli
12	700.8	93.8	750	US-09-847-513a-40	Sequence 40, Appli
13	697.6	93.4	750	US-09-847-513a-12	Sequence 12, Appli
14	696	93.2	750	US-09-847-513a-46	Sequence 46, Appli
15	694.4	93.0	750	US-09-847-513a-8	Sequence 8, Appli

16	694.4	93.0	750	US-09-847-513a-14	Sequence 14, Appli
17	694.4	93.0	750	US-09-847-513a-36	Sequence 36, Appli
18	689.6	92.3	750	US-09-847-513a-24	Sequence 24, Appli
19	688	92.1	750	US-09-847-513a-26	Sequence 26, Appli
20	459	61.4	753	US-09-847-513a-18	Sequence 18, Appli
21	459	61.4	753	US-09-847-513a-62	Sequence 62, Appli
22	457.4	61.2	753	US-09-847-513a-48	Sequence 48, Appli
23	457.4	61.2	753	US-09-847-513a-56	Sequence 56, Appli
24	455.8	61.0	753	US-09-847-513a-22	Sequence 22, Appli
25	455.8	61.0	753	US-09-847-513a-52	Sequence 52, Appli
26	454.2	60.8	753	US-09-847-513a-54	Sequence 54, Appli
27	452.6	60.6	753	US-09-847-513a-20	Sequence 20, Appli
28	452.6	60.6	753	US-09-847-513a-50	Sequence 50, Appli
29	452.6	60.6	753	US-09-847-513a-58	Sequence 58, Appli
30	452.6	60.6	753	US-09-847-513a-64	Sequence 64, Appli
31	451	60.4	753	US-09-847-513a-60	Sequence 60, Appli
32	449.4	60.2	753	US-09-847-513a-16	Sequence 16, Appli
33	48.2	6.3	7758	US-10-311-455-1076	Sequence 1, Appli
34	47.4	6.1	99116	US-10-311-455-1076	Sequence 1, Appli
35	45.4	6.1	3673778	US-10-311-455-1076	Sequence 1, Appli
36	45.2	6.1	3673778	US-10-311-455-1076	Sequence 1, Appli
37	45	6.0	11155	US-10-311-455-578	Sequence 578, Appli
38	45	6.0	12138	US-10-311-455-1916	Sequence 1916, Appli
39	45	6.0	12138	US-10-240-453-210	Sequence 210, Appli
40	44.8	6.0	8297	US-10-311-455-2047	Sequence 2047, Appli
41	44.6	6.0	6522	US-10-311-455-1023	Sequence 1023, Appli
42	44.6	6.0	9652	US-10-311-455-882	Sequence 882, Appli
43	44.4	5.9	14708	US-10-311-455-2217	Sequence 2217, Appli
44	44.4	5.9	14708	US-10-240-453-323	Sequence 323, Appli
45	44.4	5.9	14708	US-10-239-676-221	Sequence 221, Appli

#### ALIGNMENTS

RESULT 1  
US-09-847-513a-6  
Sequence 6, Application US/09847513a  
Publication No. US20030104375A1  
GENERAL INFORMATION:  
APPLICANT: MBARI  
APPLICANT: Delong, Edward  
APPLICANT: Beja, Oded  
TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin  
FILE REFERENCE: MBARI-101  
CURRENT FILING DATE: 2001-05-01  
PRIOR APPLICATION NUMBER: 60/201,602  
PRIOR FILING DATE: 2000-05-03  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 6  
LENGTH: 747  
TYPE: DNA  
ORGANISM: Naturally occurring gamma proteobacterium  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(747)  
OTHER INFORMATION: Native proteorhodopsin DNA sequence from BAC clone 31A08  
PUBLICATION INFORMATION:  
AUTHORS: Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P.,  
AUTHORS: Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N.  
AUTHORS: Delong, E.F.  
TITLE: Bacterial rhodopsin: evidence for a new type of phototrophy in the sea  
JOURNAL: Science  
VOLUME: 283  
ISSN: 5486  
PAGES: 1902-1906  
DATE: 2000-09-15  
DATABASE ACCESSION NUMBER: AAG10475  
DATABASE ENTRY DATE: 2000-06-15  
RELEVANT RESIDUES: (1)..(747)  
US-09-847-513a-6

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 15:20:39 / Search time 2081 Seconds  
(without alignments)  
8724.384 Million cell updates/sec

Title: US-09-847-513a-6

Perfect score: 747  
Sequence: 1 atgaattactatgcatact.....ttaaagaattcctaactgc 747

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_estom:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	63.2	8.5	1101	29	AL078714 Drosophila
2	58.2	7.8	1101	29	AL063072 Drosophila
3	56.4	7.6	712	13	BX416727 BX416727
4	52.6	7.0	1101	29	AL068607 Drosophila

C	5	51.8	6.9	610	12	BX34069	BX34069
6	51.8	6.9	1180	13 <td>BX436369 <td>BX436369 <td>BX436369</td> </td></td>	BX436369 <td>BX436369 <td>BX436369</td> </td>	BX436369 <td>BX436369</td>	BX436369
7	51.4	6.9	885	13 <td>BX425603 <td>BX425603 <td>BX425603</td> </td></td>	BX425603 <td>BX425603 <td>BX425603</td> </td>	BX425603 <td>BX425603</td>	BX425603
8	51.2	6.9	1101	29 <td>CNS00396 <td>CNS00396 <td>AL063921 Drosophila</td> </td></td>	CNS00396 <td>CNS00396 <td>AL063921 Drosophila</td> </td>	CNS00396 <td>AL063921 Drosophila</td>	AL063921 Drosophila
9	51	6.8	1039	13 <td>BX378025 <td>BX378025 <td>BX378025</td> </td></td>	BX378025 <td>BX378025 <td>BX378025</td> </td>	BX378025 <td>BX378025</td>	BX378025
10	50.8	6.8	990	29 <td>CNS00601 <td>CNS00601 <td>AL065624 Drosophila</td> </td></td>	CNS00601 <td>CNS00601 <td>AL065624 Drosophila</td> </td>	CNS00601 <td>AL065624 Drosophila</td>	AL065624 Drosophila
11	50.4	6.7	1101	29 <td>CNS00788 <td>CNS00788 <td>AL098379 Drosophila</td> </td></td>	CNS00788 <td>CNS00788 <td>AL098379 Drosophila</td> </td>	CNS00788 <td>AL098379 Drosophila</td>	AL098379 Drosophila
12	50.2	6.7	964	29 <td>CNS00808 <td>CNS00808 <td>AL098379 Drosophila</td> </td></td>	CNS00808 <td>CNS00808 <td>AL098379 Drosophila</td> </td>	CNS00808 <td>AL098379 Drosophila</td>	AL098379 Drosophila
13	49.4	6.6	354	28 <td>AQ090139 <td>AQ090139 <td>AL090139 GSSTC0729</td> </td></td>	AQ090139 <td>AQ090139 <td>AL090139 GSSTC0729</td> </td>	AQ090139 <td>AL090139 GSSTC0729</td>	AL090139 GSSTC0729
14	49	6.6	818	29 <td>BZ229437 <td>BZ229437 <td>BZ229437 CH230-401</td> </td></td>	BZ229437 <td>BZ229437 <td>BZ229437 CH230-401</td> </td>	BZ229437 <td>BZ229437 CH230-401</td>	BZ229437 CH230-401
15	49	6.6	909	13 <td>BX391298 <td>BX391298 <td>BX391298</td> </td></td>	BX391298 <td>BX391298 <td>BX391298</td> </td>	BX391298 <td>BX391298</td>	BX391298
16	49	6.6	1124	13 <td>BX436282 <td>BX436282 <td>BX436282</td> </td></td>	BX436282 <td>BX436282 <td>BX436282</td> </td>	BX436282 <td>BX436282</td>	BX436282
17	48.8	6.5	1097	14 <td>CD048504 <td>CD048504 <td>CD048504 AGENCOURT</td> </td></td>	CD048504 <td>CD048504 <td>CD048504 AGENCOURT</td> </td>	CD048504 <td>CD048504 AGENCOURT</td>	CD048504 AGENCOURT
18	48.8	6.5	1102	9 <td>AL572008 <td>AL572008 <td>AL572008</td> </td></td>	AL572008 <td>AL572008 <td>AL572008</td> </td>	AL572008 <td>AL572008</td>	AL572008
19	48.8	6.5	1201	13 <td>BX403984 <td>BX403984 <td>BX403984</td> </td></td>	BX403984 <td>BX403984 <td>BX403984</td> </td>	BX403984 <td>BX403984</td>	BX403984
20	48.6	6.5	1084	29 <td>CC220905 <td>CC220905 <td>CC220905 CH261-71A</td> </td></td>	CC220905 <td>CC220905 <td>CC220905 CH261-71A</td> </td>	CC220905 <td>CC220905 CH261-71A</td>	CC220905 CH261-71A
21	48.2	6.5	699	12 <td>BM630395 <td>BM630395 <td>BM630395 170006875</td> </td></td>	BM630395 <td>BM630395 <td>BM630395 170006875</td> </td>	BM630395 <td>BM630395 170006875</td>	BM630395 170006875
22	48.2	6.4	423	9 <td>AU034896 <td>AU034896 <td>AU034896 AU034896</td> </td></td>	AU034896 <td>AU034896 <td>AU034896 AU034896</td> </td>	AU034896 <td>AU034896 AU034896</td>	AU034896 AU034896
23	48	6.4	854	29 <td>CNS012CM <td>CNS012CM <td>AL101392 Drosophila</td> </td></td>	CNS012CM <td>CNS012CM <td>AL101392 Drosophila</td> </td>	CNS012CM <td>AL101392 Drosophila</td>	AL101392 Drosophila
24	47.6	6.4	966	29 <td>BZ695402 <td>BZ695402 <td>BZ695402 SP_Ba006</td> </td></td>	BZ695402 <td>BZ695402 <td>BZ695402 SP_Ba006</td> </td>	BZ695402 <td>BZ695402 SP_Ba006</td>	BZ695402 SP_Ba006
25	47.6	6.4	1198	12 <td>BM452445 <td>BM452445 <td>BM452445 AGENCOURT</td> </td></td>	BM452445 <td>BM452445 <td>BM452445 AGENCOURT</td> </td>	BM452445 <td>BM452445 AGENCOURT</td>	BM452445 AGENCOURT
26	47.4	6.3	690	29 <td>CNS0088X <td>CNS0088X <td>AL051480 Drosophila</td> </td></td>	CNS0088X <td>CNS0088X <td>AL051480 Drosophila</td> </td>	CNS0088X <td>AL051480 Drosophila</td>	AL051480 Drosophila
27	47.4	6.3	767	29 <td>CNS00A0X <td>CNS00A0X <td>AL055924 Drosophila</td> </td></td>	CNS00A0X <td>CNS00A0X <td>AL055924 Drosophila</td> </td>	CNS00A0X <td>AL055924 Drosophila</td>	AL055924 Drosophila
28	47.2	6.3	956	29 <td>CNS00LCP <td>CNS00LCP <td>AL067950 Drosophila</td> </td></td>	CNS00LCP <td>CNS00LCP <td>AL067950 Drosophila</td> </td>	CNS00LCP <td>AL067950 Drosophila</td>	AL067950 Drosophila
29	47.2	6.3	1098	13 <td>BX377526 <td>BX377526 <td>BX377526 BX377526</td> </td></td>	BX377526 <td>BX377526 <td>BX377526 BX377526</td> </td>	BX377526 <td>BX377526 BX377526</td>	BX377526 BX377526
30	47.2	6.3	1101	29 <td>CNS0170Y <td>CNS0170Y <td>AL108385 Drosophila</td> </td></td>	CNS0170Y <td>CNS0170Y <td>AL108385 Drosophila</td> </td>	CNS0170Y <td>AL108385 Drosophila</td>	AL108385 Drosophila
31	47	6.3	797	12 <td>BX324597 <td>BX324597 <td>BX324597</td> </td></td>	BX324597 <td>BX324597 <td>BX324597</td> </td>	BX324597 <td>BX324597</td>	BX324597
32	47	6.3	1007	29 <td>CNS06X9S <td>CNS06X9S <td>AL419462 T3 end of</td> </td></td>	CNS06X9S <td>CNS06X9S <td>AL419462 T3 end of</td> </td>	CNS06X9S <td>AL419462 T3 end of</td>	AL419462 T3 end of
33	46.8	6.3	527	9 <td>AU271052 <td>AU271052 <td>AU271052 AU271052</td> </td></td>	AU271052 <td>AU271052 <td>AU271052 AU271052</td> </td>	AU271052 <td>AU271052 AU271052</td>	AU271052 AU271052
34	46.8	6.3	548	9 <td>AU271053 <td>AU271053 <td>AU271053 AU271053</td> </td></td>	AU271053 <td>AU271053 <td>AU271053 AU271053</td> </td>	AU271053 <td>AU271053 AU271053</td>	AU271053 AU271053
35	46.8	6.3	992	29 <td>CNS002SH <td>CNS002SH <td>AL063310 Drosophila</td> </td></td>	CNS002SH <td>CNS002SH <td>AL063310 Drosophila</td> </td>	CNS002SH <td>AL063310 Drosophila</td>	AL063310 Drosophila
36	46.8	6.3	1001	29 <td>CNS015SH <td>CNS015SH <td>AL105023 Drosophila</td> </td></td>	CNS015SH <td>CNS015SH <td>AL105023 Drosophila</td> </td>	CNS015SH <td>AL105023 Drosophila</td>	AL105023 Drosophila
37	46.6	6.2	1200	9 <td>AL548181 <td>AL548181 <td>AL548181</td> </td></td>	AL548181 <td>AL548181 <td>AL548181</td> </td>	AL548181 <td>AL548181</td>	AL548181
38	46.2	6.2	711	28 <td>BZ098560 <td>BZ098560 <td>BZ098560 CH230-235</td> </td></td>	BZ098560 <td>BZ098560 <td>BZ098560 CH230-235</td> </td>	BZ098560 <td>BZ098560 CH230-235</td>	BZ098560 CH230-235
39	46.2	6.2	1163	13 <td>BX415221 <td>BX415221 <td>BX415221</td> </td></td>	BX415221 <td>BX415221 <td>BX415221</td> </td>	BX415221 <td>BX415221</td>	BX415221
40	46.2	6.2	1167	29 <td>CNS07360 <td>CNS07360 <td>AL427102 clone BAO</td> </td></td>	CNS07360 <td>CNS07360 <td>AL427102 clone BAO</td> </td>	CNS07360 <td>AL427102 clone BAO</td>	AL427102 clone BAO
41	46.2	6.2	1175	12 <td>B1872945 <td>B1872945 <td>B1872945 603398115</td> </td></td>	B1872945 <td>B1872945 <td>B1872945 603398115</td> </td>	B1872945 <td>B1872945 603398115</td>	B1872945 603398115
42	46.2	6.2	1201	13 <td>BX336382 <td>BX336382 <td>BX336382</td> </td></td>	BX336382 <td>BX336382 <td>BX336382</td> </td>	BX336382 <td>BX336382</td>	BX336382
43	46	6.1	1201	13 <td>BX417935 <td>BX417935 <td>BX417935</td> </td></td>	BX417935 <td>BX417935 <td>BX417935</td> </td>	BX417935 <td>BX417935</td>	BX417935
44	45.8	6.1	884	29 <td>CNS0129A <td>CNS0129A <td>AL101272 Drosophila</td> </td></td>	CNS0129A <td>CNS0129A <td>AL101272 Drosophila</td> </td>	CNS0129A <td>AL101272 Drosophila</td>	AL101272 Drosophila
45	45.6	6.1	423	28 <td>AZ392278 <td>AZ392278 <td>AZ392278 IM0154M12</td> </td></td>	AZ392278 <td>AZ392278 <td>AZ392278 IM0154M12</td> </td>	AZ392278 <td>AZ392278 IM0154M12</td>	AZ392278 IM0154M12

## ALIGNMENTS

RESULT 1  
CNS00L72  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TERT end of BAC:  
BACK48P19 of RPLT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
AL078714  
VERSION  
AL078714.1 GI:5102004  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Preygotha;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuo Osega and  
Aaron Mammor in Pieter de Jong's laboratory in the Department of

## COMMENT

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2003, 04:11:46 ; Search time 382.939 Seconds  
(Without alignments)  
4410.621 Million cell updates/sec

Title: US-09-847-513a-4  
Perfect score: 750  
Sequence: 1 atgggaataattactatgcat.....ttaagaattcttactatgct 750

Scoring table: IDENTITY\_NNC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Listing first 100 summaries

Database : N.Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750	100.0	750	24	ABK24125
2	745.2	99.4	750	24	ABK24139
3	744	99.2	747	24	ABK24126
4	744	99.2	105184	24	ABK24132
5	743.6	99.1	750	24	ABK24128
6	734	97.9	750	24	ABK24142
7	734	97.9	750	24	ABK24145
8	719.6	95.9	750	24	ABK24140
9	711.6	94.9	750	24	ABK24138

10	711.6	94.9	750	24	ABK24144	DNA encoding Prote
11	710	94.7	750	24	ABK24137	DNA encoding Prote
12	706.8	94.2	750	24	ABK24143	DNA encoding Prote
13	703.6	93.8	750	24	ABK24129	DNA encoding Prote
14	702	93.6	750	24	ABK24116	DNA encoding Prote
15	700.4	93.4	750	24	ABK24117	DNA encoding Prote
16	700.4	93.4	750	24	ABK24130	DNA encoding Prote
17	700.4	93.4	750	24	ABK24141	DNA encoding Prote
18	695.6	92.7	750	24	ABK24135	DNA encoding Prote
19	694	92.5	750	24	ABK24136	DNA encoding Prote
20	465	62.0	753	24	ABK24132	DNA encoding Prote
21	465	62.0	753	24	ABK24134	DNA encoding Prote
22	463.4	61.8	753	24	ABK24147	DNA encoding Prote
23	463.4	61.8	753	24	ABK24151	DNA encoding Prote
24	461.8	61.6	753	24	ABK24134	DNA encoding Prote
25	461.8	61.6	753	24	ABK24150	DNA encoding Prote
26	460.2	61.4	753	24	ABK24133	DNA encoding Prote
27	458.0	61.1	753	24	ABK24148	DNA encoding Prote
28	458.0	61.1	753	24	ABK24152	DNA encoding Prote
29	458.6	61.1	753	24	ABK24155	DNA encoding Prote
30	457	60.9	753	24	ABK24153	DNA encoding Prote
31	455.4	60.7	753	24	ABK24131	DNA encoding Prote
32	455.4	60.7	753	24	ABK24131	DNA encoding Prote
33	48.6	6.5	6767	22	AA546608	DNA encoding Prote
34	48.6	6.5	7131	24	ABL70427	DNA encoding Prote
35	48.6	6.5	7131	24	AA561360	DNA encoding Prote
36	48.6	6.5	7131	24	ABK31450	DNA encoding Prote
37	47.4	6.3	7758	24	ABL33103	DNA encoding Prote
38	45.6	6.1	6317	24	ABL49312	DNA encoding Prote
39	45.6	6.1	6317	24	ABL32409	DNA encoding Prote
40	45	6.0	11155	24	ABL32605	DNA encoding Prote
41	45	6.0	12138	24	ABL33943	DNA encoding Prote
42	45	6.0	12138	24	ABK28336	DNA encoding Prote
43	44.8	6.0	6419	24	ABL32266	DNA encoding Prote
44	44.8	6.0	8297	24	ABL34074	DNA encoding Prote
45	44.8	6.0	56153	22	AA546794	DNA encoding Prote
46	44.6	5.9	390	24	ABA97217	DNA encoding Prote
47	44.6	5.9	6522	24	ABL33050	DNA encoding Prote
48	44.6	5.9	9652	24	ABL32909	DNA encoding Prote
49	44.6	5.9	12639	24	ABN80107	DNA encoding Prote
50	44.4	5.9	3584	23	ABL07253	DNA encoding Prote
51	44.4	5.9	6030	22	AA545456	DNA encoding Prote
52	44.4	5.9	6030	24	AA561348	DNA encoding Prote
53	44.4	5.9	6030	24	AA5613312	DNA encoding Prote
54	44.4	5.9	8776	24	ABK28312	DNA encoding Prote
55	44.4	5.9	13531	23	ABL07252	DNA encoding Prote
56	44.2	5.9	12000	24	ABK34014	DNA encoding Prote
57	44.2	5.9	14362	24	ABQ67047	DNA encoding Prote
58	44.2	5.9	14708	22	AA545513	DNA encoding Prote
59	44.2	5.9	14708	22	AA546773	DNA encoding Prote
60	44.2	5.9	14708	24	ABL92324	DNA encoding Prote
61	44.2	5.9	14708	24	ABL34244	DNA encoding Prote
62	44.2	5.9	14708	24	ABK28449	DNA encoding Prote
63	43.8	5.8	5912	24	ABL32603	DNA encoding Prote
64	43.8	5.8	6292	22	AA546735	DNA encoding Prote
65	43.2	5.8	7001	24	ABK33919	DNA encoding Prote
66	43	5.7	4122	23	ABL28439	DNA encoding Prote
67	43	5.7	5273	24	ABL32875	DNA encoding Prote
68	43	5.7	6535	24	ABL32936	DNA encoding Prote
69	43	5.7	6708	23	ABL28438	DNA encoding Prote
70	43	5.7	10266	17	AA173007	DNA encoding Prote
71	43	5.7	11178	24	ABL70507	DNA encoding Prote
72	43	5.7	11178	24	AA561059	DNA encoding Prote
73	43	5.7	11178	24	ABK31172	DNA encoding Prote
74	43	5.7	14704	13	AAQ20685	DNA encoding Prote
75	42.8	5.7	875	22	AA546384	DNA encoding Prote
76	42.6	5.7	5649	22	AA546384	DNA encoding Prote
77	42.6	5.7	5649	24	ABK40008	DNA encoding Prote
78	42.6	5.7	5649	24	ABL32889	DNA encoding Prote
79	42.6	5.7	6283	24	ABL32088	DNA encoding Prote
80	42.4	5.6	16688	22	AA546556	DNA encoding Prote
81	42.2	5.6	1448	24	ABQ14794	DNA encoding Prote
82	42.2	5.6	1448	24	ABQ14795	DNA encoding Prote

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2003, 05:27:32 ; Search time 2805.73 Seconds

(Without alignments)  
7779.478 Million cell updates/sec

Title: US-09-847-513a-4

Sequence: 1 atgggaattattactgac.....ttaagaattcttaattgct 750

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

```

GenBank:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_ph:*
7: gb_pl:*
8: gb_pl:*
9: gb_pl:*
10: gb_pl:*
11: gb_pl:*
12: gb_pl:*
13: gb_pl:*
14: gb_pl:*
15: gb_pl:*
16: gb_pl:*
17: gb_pl:*
18: gb_pl:*
19: gb_pl:*
20: gb_pl:*
21: gb_pl:*
22: gb_pl:*
23: gb_pl:*
24: gb_pl:*
25: gb_pl:*
26: gb_pl:*
27: gb_pl:*
28: gb_pl:*
29: gb_pl:*
30: gb_pl:*
31: gb_pl:*
32: gb_pl:*
33: gb_pl:*
34: gb_pl:*
35: gb_pl:*
36: gb_pl:*
37: gb_pl:*
38: gb_pl:*
39: gb_pl:*
40: gb_pl:*
41: gb_pl:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	744	99.2	105184	1	AF279106
2	740.2	98.7	748	1	AF349987
3	734	97.9	753	1	AF349990
4	734	97.9	753	1	AF349993
5	719.6	95.9	753	1	AF349988
6	711.6	94.9	753	1	AF349986
7	711.6	94.9	753	1	AF349992
8	710	94.7	753	1	AF349985
9	706.8	93.8	753	1	AF349991
10	703.6	93.6	757	1	AF349977
11	702	93.6	753	1	AF349994
12	700.4	93.4	753	1	AF349978
13	700.4	93.4	757	1	AF349976
14	695.6	92.7	753	1	AF349983
15	695.4	92.7	748	1	AF349989
16	694	92.5	753	1	AF349984
17	465	62.0	756	1	AF350002
18	465	62.0	756	1	AF350000
19	463.4	61.8	756	1	AF349995
20	463.4	61.8	756	1	AF349980
21	461.8	61.6	756	1	AF349998
22	460.2	61.4	756	1	AF349982
23	458.6	61.1	756	1	AF349997
24	458.6	61.1	756	1	AF349999
25	458.6	61.1	756	1	AF349981
26	458.6	61.1	756	1	AF349996
27	458.6	61.1	756	1	AF350000
28	457	60.9	756	1	AF350003
29	455.4	60.7	756	1	AF350001
30	57.2	7.6	236542	2	AF349979
31	54.4	7.3	169163	2	AC096320
32	54	7.2	84472	2	AC115666
33	54	7.2	176351	2	AC096684
34	53.6	7.1	163034	2	AC117007
35	53.4	7.1	108476	2	AC099432
36	52.8	7.0	138156	2	AC094507
37	52.6	7.0	131346	2	AC111482
38	52.6	7.0	175699	2	AC119558
39	52.6	7.0	175699	2	AC102515
40	52.6	7.0	201097	2	AC129136
41	52.6	7.0	207364	2	AC112091
42	52	6.9	125026	10	AL671857
43	52	6.9	165337	2	AC096212
44	51.6	6.9	141008	2	AC126897
45	51.6	6.9	264522	2	AC090437
46	51.4	6.9	7921	3	AF153362
47	51.4	6.9	138825	2	AC125755
48	51.2	6.8	111338	2	AC096844
49	51	6.8	97683	2	AC116548
50	50.4	6.7	144225	2	AC118320
51	50.4	6.7	151705	2	AC117900
52	50.2	6.7	173540	2	AC119697
53	50.2	6.7	196650	2	AC111217
54	49.8	6.6	174599	2	AC106523
55	49.6	6.6	71553	2	AC096997
56	49.4	6.6	156533	2	AC117070
57	49	6.5	97089	2	AC129683
58	49	6.5	101739	2	AC114251
59	49	6.5	161799	2	AC098991
60	49	6.5	197627	2	AC112735
61	49	6.5	211027	2	AC130161
62	48.8	6.5	22526	2	AC129738
63	48.8	6.5	79981	2	AC126315
64	48.8	6.5	127573	2	AC097158
65	48.8	6.5	165318	2	AC126747

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2003, 06:04:58 ; Search time 79.3898 Seconds

(without alignments)

2897.193 Million cell updates/sec

Title: US-09-847-513A-4

Perfect score: 750

Sequence: 1. atggaataattactgct.....ttaagaattcttaagtct 750

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	42	5.6	9793 1	US-08-470-202-56
2	42	5.6	9793 1	US-08-471-770-56
3	42	5.6	9793 2	US-08-468-059-56
4	42	5.6	9793 4	US-09-109-916-56
5	39.8	5.3	774 4	US-09-134-001C-2190
6	38.8	5.2	1395 1	US-07-991-867B-25
7	38.8	5.2	1395 1	US-08-107-755A-25
8	38.8	5.2	1395 2	US-08-544-332-25
9	38.8	5.2	1395 4	US-09-370-861A-25
10	38.8	5.2	6768 1	US-08-107-755A-1
11	38.8	5.2	8457 1	US-07-991-867B-1
12	38.8	5.2	8457 2	US-08-544-332-1
13	38.8	5.2	8457 4	US-09-370-861A-1
14	38.4	5.1	5511 3	US-08-928-361B-2
15	38.4	5.1	7334 3	US-08-928-361B-1
16	38	5.0	1284 4	US-09-134-001C-2306
17	37.2	5.0	397 3	US-09-253-681-3
18	36.8	4.9	1086 1	US-08-415-751-27
19	36.8	4.9	1086 3	US-08-415-751-28
20	36.8	4.9	5163 3	US-08-700-651-1
21	36.8	4.9	5163 3	US-08-928-361B-4
22	36.8	4.9	5318 3	US-08-700-651-2
23	36.8	4.9	5318 3	US-08-928-361B-3
24	36.6	4.9	1447 4	US-09-227-357-17
25	36.4	4.9	7218 1	US-08-232-463-14
26	36	4.8	2190 4	US-09-625-188-19
27	35.6	4.7	663 4	US-08-998-416-187

28	35.6	4.7	719	4	US-08-998-416-1138	Sequence 1138, App
29	35.6	4.7	856	4	US-08-998-416-289	Sequence 289, App
30	35.4	4.7	11091	4	US-09-1134-001C-2243	Sequence 2243, App
31	35.2	4.7	2824	4	US-07-757-022B-13	Sequence 13, App
32	35.2	4.7	3066	4	US-07-757-022B-83	Sequence 83, App
33	35.2	4.7	3117	4	US-07-757-022B-73	Sequence 73, App
34	35.2	4.7	3148	4	US-07-757-022B-57	Sequence 57, App
35	35.2	4.7	3292	1	US-07-814-964-12	Sequence 12, App
36	35.2	4.7	3292	1	US-08-258-442-12	Sequence 12, App
37	35.2	4.7	3292	1	US-08-328-809-7	Sequence 7, App
38	35.2	4.7	3292	5	PCT-US92-11107-12	Sequence 12, App
39	35.2	4.7	3376	3	US-08-320-559-29	Sequence 29, App
40	35.2	4.7	3376	5	US-08-545-860D-29	Sequence 29, App
41	35.2	4.7	3376	5	PCT-US94-04496-29	Sequence 29, App
42	35.2	4.7	3420	4	US-07-757-022B-103	Sequence 103, App
43	35.2	4.7	3813	4	US-07-757-022B-43	Sequence 43, App
44	35.2	4.7	3831	4	US-08-961-527-291	Sequence 291, App
45	35.2	4.7	3942	4	US-07-757-022B-141	Sequence 141, App
46	35.2	4.7	3942	4	US-07-757-022B-45	Sequence 45, App
47	35.2	4.7	3942	4	US-07-757-022B-39	Sequence 39, App
48	35.2	4.7	3963	4	US-07-757-022B-59	Sequence 59, App
49	35.2	4.7	4065	4	US-07-757-022B-47	Sequence 47, App
50	35.2	4.7	4065	4	US-07-757-022B-39	Sequence 39, App
51	35.2	4.7	4092	4	US-07-757-022B-51	Sequence 51, App
52	35.2	4.7	4215	4	US-07-757-022B-61	Sequence 61, App
53	35.2	4.7	4215	4	US-07-757-022B-61	Sequence 61, App
54	35.2	4.7	5008	4	US-07-757-022B-1	Sequence 1, App
55	35.2	4.7	6211	4	US-08-961-527-8	Sequence 8, App
56	34.8	4.6	234	1	US-08-469-802B-3	Sequence 3, App
57	34.8	4.6	234	2	US-08-267-803B-3	Sequence 3, App
58	34.6	4.6	468	4	US-09-134-001C-2729	Sequence 2729, App
59	34.6	4.6	854	3	US-08-913-842-16	Sequence 16, App
60	34.4	4.6	2238	2	US-08-913-842-2	Sequence 2, App
61	34.2	4.6	1785	4	US-09-468-872-1	Sequence 1, App
62	34.2	4.6	3366	4	US-08-961-527-136	Sequence 136, App
63	34	4.5	198	5	PCT-US95-10668-3	Sequence 3, App
64	34	4.5	198	5	PCT-US95-10668-4	Sequence 4, App
65	33.8	4.5	203	4	US-09-043-303-7	Sequence 7, App
66	33.8	4.5	397	4	US-08-920-812-6	Sequence 6, App
67	33.8	4.5	8654	1	US-08-920-812-6	Sequence 6, App
68	33.8	4.5	8654	1	US-08-920-812-6	Sequence 6, App
69	33.8	4.5	8654	1	US-08-921-177-6	Sequence 6, App
70	33.8	4.5	8654	1	US-08-362-577C-6	Sequence 6, App
71	33.8	4.5	8654	2	US-08-920-812-6	Sequence 6, App
72	33.6	4.5	642	4	US-08-858-207A-48	Sequence 48, App
73	33.6	4.5	19124	2	US-08-487-826B-13	Sequence 13, App
74	33.4	4.5	543	6	5273901-6	Sequence 6, App
75	33.4	4.5	3280	1	US-08-259-000-4	Sequence 4, App
76	33.4	4.5	3280	1	US-08-729-767-6	Sequence 6, App
77	33.2	4.4	533	6	5482709-5	Sequence 5, App
78	33.2	4.4	1212	3	US-09-046-578-1	Sequence 1, App
79	33.2	4.4	6709	4	US-09-268-601-3	Sequence 3, App
80	33.2	4.4	7604	4	US-09-268-992-7	Sequence 7, App
81	33.2	4.4	7604	4	US-09-657-474-7	Sequence 7, App
82	33.2	4.4	1236	4	US-09-134-001C-462	Sequence 462, App
83	33	4.4	5181	1	US-08-257-073-10	Sequence 10, App
84	32.8	4.4	1167	4	US-09-308-003-6	Sequence 6, App
85	32.8	4.4	1858	2	US-08-909-965C-11	Sequence 11, App
86	32.8	4.4	14440	4	US-08-961-527-128	Sequence 128, App
87	32.8	4.4	87563	4	US-09-453-702B-57	Sequence 57, App
88	32.6	4.3	198	5	PCT-US95-10668-2	Sequence 2, App
89	32.6	4.3	198	5	PCT-US95-10668-2	Sequence 2, App
90	32.6	4.3	511	4	US-09-175-658B-22	Sequence 22, App
91	32.6	4.3	652	4	US-09-036-574-9	Sequence 9, App
92	32.6	4.3	838	4	US-09-036-574-5	Sequence 5, App
93	32.6	4.3	1621	1	US-08-480-604A-27	Sequence 27, App
94	32.6	4.3	3891	1	US-08-405-496A-27	Sequence 27, App
95	32.6	4.3	3891	2	US-08-915-136-27	Sequence 27, App
96	32.6	4.3	3891	2	US-08-915-136-27	Sequence 27, App
97	32.6	4.3	6492	4	US-08-961-527-188	Sequence 188, App
98	32.6	4.3	38844	4	US-09-734-675-3	Sequence 3, App
99	32.4	4.3	3666	2	US-08-682-517-13	Sequence 13, App
100	32.4	4.3	3666	2	US-08-682-517-14	Sequence 14, App

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2003, 06:12:17 ; Search time 229.763 Seconds

(without alignments)  
4310.289 Million cell updates/sec

Title: **us-09-847-513a-4**

Perfect score: 750  
Sequence: 1 atggcgaattactatgcgtat.....ttaagaattcttaagtct 750

Scoring table: IDENTITY\_NUC

Gap 10.0 , Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications, NA:\*

```

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCOT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCOTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.4	5.9	6030	9	US-10-239-676-164
2	44.2	5.9	14708	9	US-10-239-676-221
3	43.8	5.8	1236	9	US-10-077-584-3
4	42.2	5.6	2614	9	US-09-822-846-491
5	42.2	5.6	6604	10	US-09-880-107-1748
6	42	5.6	9793	9	US-09-886-156-56
7	42	5.6	9793	9	US-09-886-150-56
8	42	5.6	9793	9	US-09-886-149-56
9	42	5.6	9793	9	US-09-886-159-56
10	39.6	5.3	5908	9	US-10-239-676-93
11	39.2	5.2	6730	7	US-08-781-986A-51
12	38.8	5.2	7906	9	US-10-239-676-97
13	38.4	5.1	426	10	US-09-960-352-8406
14	38	5.1	5986	9	US-10-239-676-145
15	37.8	5.0	17419	9	US-10-239-676-99
16	37.6	5.0	2203	10	US-09-925-302-113
17	37.4	5.0	4239	10	US-09-070-927A-450
18	37	4.9	818	9	US-09-764-891-256
19	37	4.9	6053	9	US-10-239-676-76

20	36.8	4.9	2000	10	US-09-887-576-58	Sequence 58, Appl
21	36.8	4.9	143306	10	US-09-729-920-3	Sequence 3, Appl1
22	36.6	4.9	1447	9	US-09-983-802-17	Sequence 17, Appl
23	36.6	4.9	17421	9	US-10-239-676-53	Sequence 53, Appl
24	36.4	4.9	5976	9	US-10-239-676-17	Sequence 17, Appl
25	36.2	4.8	12465	9	US-10-239-676-31	Sequence 31, Appl
26	36.2	4.8	15649	9	US-10-239-676-104	Sequence 104, Appl
27	36	4.8	567	10	US-09-070-927A-664	Sequence 664, Appl
28	36	4.8	1194	10	US-09-815-242-8601	Sequence 8601, Ap
29	35.8	4.8	830	10	US-09-864-761-19531	Sequence 19531, A
30	35.8	4.8	5898	9	US-10-239-676-179	Sequence 179, App
31	35.6	4.7	1310	9	US-09-849-243-13	Sequence 13, Appl
32	35.6	4.7	3263	9	US-09-849-243-15	Sequence 15, Appl
33	35.6	4.7	4286	9	US-09-849-243-14	Sequence 14, Appl
34	35.6	4.7	8996	9	US-10-239-676-211	Sequence 211, App
35	35.4	4.7	425	9	US-09-918-995-8119	Sequence 8119, Ap
36	35.4	4.7	11036	9	US-10-239-676-117	Sequence 117, App
37	35.4	4.7	17142	9	US-10-239-676-205	Sequence 205, App
38	35.4	4.7	17846	9	US-08-781-986A-109	Sequence 109, App
39	35.2	4.7	488	9	US-09-918-995-19918	Sequence 19918, A
40	35.2	4.7	553	10	US-09-920-300A-1461	Sequence 1461, Ap
41	35.2	4.7	553	12	US-10-033-528-1461	Sequence 1461, Ap
42	35.2	4.7	1329	10	US-09-815-242-9139	Sequence 9139, Ap
43	35.2	4.7	1329	10	US-09-815-242-9139	Sequence 9139, Ap
44	35.2	4.7	2824	12	US-10-124-557-13	Sequence 13, Appl
45	35.2	4.7	3066	12	US-10-124-557-83	Sequence 83, Appl
46	35.2	4.7	3117	12	US-10-124-557-73	Sequence 73, Appl
47	35.2	4.7	3148	12	US-10-124-557-57	Sequence 57, Appl
48	35.2	4.7	3420	12	US-10-124-557-103	Sequence 103, Appl
49	35.2	4.7	3813	12	US-10-124-557-43	Sequence 43, Appl
50	35.2	4.7	3936	12	US-10-124-557-41	Sequence 41, Appl
51	35.2	4.7	3942	12	US-10-124-557-141	Sequence 141, App
52	35.2	4.7	3945	12	US-10-124-557-49	Sequence 49, Appl
53	35.2	4.7	3963	12	US-10-124-557-45	Sequence 45, Appl
54	35.2	4.7	3963	12	US-10-124-557-59	Sequence 59, Appl
55	35.2	4.7	4065	12	US-10-124-557-47	Sequence 47, Appl
56	35.2	4.7	4082	12	US-10-124-557-39	Sequence 39, Appl
57	35.2	4.7	4096	12	US-10-124-557-51	Sequence 51, Appl
58	35.2	4.7	4215	12	US-10-124-557-61	Sequence 61, Appl
59	35.2	4.7	4575	12	US-10-044-090-303	Sequence 303, Appl
60	35.2	4.7	5008	12	US-10-124-557-1	Sequence 1, Appl1
61	35.2	4.7	6665	9	US-10-239-676-4	Sequence 4, Appl1
62	35.2	4.7	10286	9	US-10-239-676-13	Sequence 13, Appl
63	35.2	4.7	15306	9	US-10-239-676-162	Sequence 162, App
64	35	4.7	2341	10	US-09-881-752A-147	Sequence 147, App
65	35	4.7	4605	10	US-09-070-927A-315	Sequence 315, App
66	35	4.7	7104	10	US-09-815-242-4580	Sequence 4580, Ap
67	35	4.7	7107	10	US-09-815-242-8291	Sequence 8291, Ap
68	35	4.7	7237	10	US-09-070-927A-239	Sequence 239, App
69	35	4.7	7434	10	US-09-815-242-4761	Sequence 4761, Ap
70	35	4.7	7437	10	US-09-815-242-8869	Sequence 8869, Ap
71	35	4.7	8155	7	US-08-781-986A-63	Sequence 63, Appl
72	35	4.7	9539	9	US-10-239-676-52	Sequence 52, Appl
73	34.8	4.6	325	10	US-09-878-574-451	Sequence 451, App
74	34.8	4.6	449	9	US-09-918-995-26938	Sequence 26938, A
75	34.8	4.6	512	10	US-09-864-761-25347	Sequence 25347, A
76	34.8	4.6	575	10	US-09-864-761-8628	Sequence 8628, Ap
77	34.8	4.6	6078	9	US-10-239-676-173	Sequence 173, App
78	34.8	4.6	6203	9	US-10-239-676-183	Sequence 183, Appl
79	34.8	4.6	6203	9	US-10-239-676-63	Sequence 63, Appl
80	34.8	4.6	6398	9	US-10-239-676-96	Sequence 96, Appl
81	34.6	4.6	319	9	US-09-803-719-2197	Sequence 2197, Ap
82	34.6	4.6	6620	9	US-10-239-676-195	Sequence 195, App
83	34.6	4.6	7195	9	US-10-239-676-30	Sequence 30, Appl
84	34.6	4.6	46537	10	US-09-933-267A-1	Sequence 1, Appl1
85	34.4	4.6	337	10	US-09-960-352-95976	Sequence 6976, Ap
86	34.4	4.6	2000	9	US-09-938-842A-2742	Sequence 2742, Ap
87	34.4	4.6	3414	9	US-09-938-842A-3356	Sequence 3356, Ap
88	34.4	4.6	3414	9	US-10-074-475-36	Sequence 36, Appl
89	34.4	4.6	6397	9	US-10-239-676-107	Sequence 107, App
90	34.4	4.6	20633	10	US-09-070-927A-276	Sequence 276, App
91	34.2	4.6	293	10	US-09-864-761-18923	Sequence 18923, A
92	34.2	4.6	459	10	US-09-864-761-2182	Sequence 2182, Ap